

#2
PATENTS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Atsushi Murashima

Serial No.: unassigned

Art Unit: unassigned

Filed: herewith

Docket: 14667

For: VOICE DETECTING METHOD AND
APPARATUS, AND MEDIUM THEREOF

Dated: May 31, 2001



Assistant Commissioner for Patents
Washington, DC 20231

CLAIM OF PRIORITY

Sir:

Applicant, in the above-identified application, hereby claims the right of priority in connection with Title 35 U.S.C. §119 and in support thereof, herewith submits a certified copy of Japanese Patent Application No. 2000-166746, filed on June 2, 2000.

Respectfully submitted,

Paul J. Esatto, Jr.
Registration No. 30,749

Scully, Scott, Murphy & Presser
400 Garden City Plaza
Garden City, NY 11530
(516) 742-4343
PJE:vjs

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

"Express Mail" Mailing Label Number: EL915257223US
Date of Deposit: May 31, 2001

I hereby certify that this correspondence is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. §1.10 on the date indicated above and is addressed to the Assistant Commissioner of Patents and Trademarks, Washington, D.C. 20231.

Dated: May 31, 2001

Janet Grossman

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

2000年 6月 2日

出 願 番 号

Application Number:

特願2000-166746

出 願 人

Applicant (s):

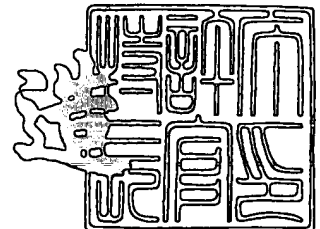
日本電気株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2001年 3月23日

特許庁長官
Commissioner,
Patent Office

及川耕



【書類名】 特許願

【整理番号】 33509760

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00

【発明者】

 【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内

 【氏名】 村島 淳

【特許出願人】

 【識別番号】 000004237

 【氏名又は名称】 日本電気株式会社

【代理人】

 【識別番号】 100079005

 【弁理士】

 【氏名又は名称】 宇高 克己

【手数料の表示】

 【予納台帳番号】 009265

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

 【包括委任状番号】 9715827

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声検出方法及び装置とその記録媒体

【特許請求の範囲】

【請求項 1】 一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出方法において、

特徴量の変動量をフィルタに入力することで得られる変動量の長時間平均を用いて、音声信号を一定時間長毎に音声区間と非音声区間とに判別することを特徴とする音声検出方法。

【請求項 2】 前記特徴量の変動量を、前記特徴量とその長時間平均とを用いて計算することを特徴とする請求項 1 に記載の音声検出方法。

【請求項 3】 請求項 1 又は請求項 2 に記載の音声検出方法によって、過去に出力された判別の結果を用いて、前記変動量の長時間平均を計算する際の前記フィルタを切り替えることを特徴とする請求項 1 又は請求項 2 に記載の音声検出方法。

【請求項 4】 過去に入力された音声信号から計算される特徴量を用いることを特徴とする請求項 1 から請求項 3 のいずれかに記載の音声検出方法。

【請求項 5】 前記特徴量として、線スペクトル周波数、全帯域エネルギー、低域エネルギー及び零交叉数のうちの少なくとも一つを用いることを特徴とする請求項 1 から請求項 4 のいずれかに記載の音声検出方法。

【請求項 6】 音声復号方法によって復号される線形予測係数から計算される線スペクトル周波数と、前記音声復号方法によって過去に出力された再生音声信号から計算される全帯域エネルギー、低域エネルギー及び零交叉数のうちの少なくとも一つを用いることを特徴とする請求項 5 に記載の音声検出方法。

【請求項 7】 一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出装置において、

音声信号から線スペクトル周波数 (LSF) を計算する LSF 計算回路と、

前記音声信号から全帯域エネルギーを計算する全帯域エネルギー計算回路と、

前記音声信号から低域エネルギーを計算する低域エネルギー計算回路と、
 前記音声信号から零交叉数を計算する零交叉数計算回路と、
 前記線スペクトル周波数の変動量（第 1 の変動量）を計算する線スペクトル周波数変動量計算部と、 前記全帯域エネルギーの変動量（第 2 の変動量）を計算する全帯域エネルギー変動量計算部と、 前記低域エネルギーの変動量（第 3 の変動量）を計算する低域エネルギー変動量計算部と、
 前記零交叉数の変動量（第 4 の変動量）を計算する零交叉数変動量計算部と、
 前記第 1 の変動量の長時間平均を計算する第 1 のフィルタと、
 前記第 2 の変動量の長時間平均を計算する第 2 のフィルタと、
 前記第 3 の変動量の長時間平均を計算する第 3 のフィルタと、
 前記第 4 の変動量の長時間平均を計算する第 4 のフィルタと
 を含むことを特徴とする音声検出装置。

【請求項 8】 一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出装置において、

音声信号から線スペクトル周波数（LSF）を計算する LSF 計算回路と、
 前記音声信号から全帯域エネルギーを計算する全帯域エネルギー計算回路と、
 前記音声信号から低域エネルギーを計算する低域エネルギー計算回路と、
 前記音声信号から零交叉数を計算する零交叉数計算回路と、
 前記線スペクトル周波数とその長時間平均との差分に基づく第 1 の変動量を計算する第 1 の変動量計算回路と、
 前記全帯域エネルギーとその長時間平均との差分に基づく第 2 の変動量を計算する第 2 の変動量計算回路と、
 前記低域エネルギーとその長時間平均との差分に基づく第 3 の変動量を計算する第 3 の変動量計算回路と、
 前記零交叉数とその長時間平均との差分に基づく第 4 の変動量を計算する第 4 の変動量計算回路と、
 前記第 1 の変動量の長時間平均を計算する第 1 のフィルタと、
 前記第 2 の変動量の長時間平均を計算する第 2 のフィルタと、

前記第 3 の変動量の長時間平均を計算する第 3 のフィルタと、
前記第 4 の変動量の長時間平均を計算する第 4 のフィルタと
を含むことを特徴とする音声検出装置。

【請求項 9】 請求項 7 又は請求項 8 に記載の音声検出装置から過去に出力された前記判別の結果を保持する第 1 の記憶回路と、

前記第 1 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 5 のフィルタと第 6 のフィルタとを切り替える第 1 の切替器と、

前記第 2 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 7 のフィルタと第 8 のフィルタとを切り替える第 2 の切替器と、

前記第 3 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 9 のフィルタと第 10 のフィルタとを切り替える第 3 の切替器と、

前記第 4 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 11 のフィルタと第 12 のフィルタとを切り替える第 4 の切替器と

を含むことを特徴とする請求項 7 または 8 に記載の音声検出装置。

【請求項 10】 過去に入力された前記音声信号から前記線スペクトル周波数と、前記全帯域エネルギーと、前記低域エネルギーと、前記零交叉数とを計算することを特徴とする請求項 7 から請求項 9 のいずれかに記載の音声検出装置。

【請求項 11】 前記特徴量として、線スペクトル周波数、全帯域エネルギー、低域エネルギーおよび零交叉数のうちの少なくとも一つを用いることを特徴とする請求項 7 から 10 のいずれかに記載の音声検出装置。

【請求項 12】 音声復号装置から過去に出力された再生音声信号を記憶保持する第 2 の記憶回路を備え、

前記第 2 の記憶回路から出力される前記再生音声信号から計算される全帯域エネルギー、低域エネルギーおよび零交叉数と、前記音声復号装置において復号される線形予測係数から計算される線スペクトル周波数とのうちの少なくとも一つ

を用いることを特徴とする請求項 7 から請求項 1 1 のいずれかに記載の音声検出装置。

【請求項 1 3】 一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出装置を構成する情報処理装置に、

(a) 前記音声信号から線スペクトル周波数 (LSF) を計算する処理と、 (b) 前記音声信号から全帯域エネルギーを計算する処理と、 (c) 前記音声信号から低域エネルギーを計算する処理と、 (d) 前記音声信号から零交叉数を計算する処理と、 (e) 前記線スペクトル周波数の変動量 (第 1 の変動量) を計算する処理と、 (f) 前記全帯域エネルギーの変動量 (第 2 の変動量) を計算する処理と、 (g) 前記低域エネルギーの変動量 (第 3 の変動量) を計算する処理と、

(h) 前記零交叉数の変動量 (第 4 の変動量) を計算する処理と、

(I) 前記第 1 の変動量の長時間平均を計算する処理と、

(j) 前記第 2 の変動量の長時間平均を計算する処理と、

(k) 前記第 3 の変動量の長時間平均を計算する処理と、

(l) 前記第 4 の変動量の長時間平均を計算する処理と、

の前記 (a) から (l) の処理を実行させるためのプログラムを記録した前記情報処理装置が読み取り可能な記録媒体。

【請求項 1 4】 一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出装置を構成する情報処理装置に、

(a) 前記音声信号から線スペクトル周波数 (LSF) を計算する処理と、

(b) 前記音声信号から全帯域エネルギーを計算する処理と、

(c) 前記音声信号から低域エネルギーを計算する処理と、

(d) 前記音声信号から零交叉数を計算する処理と、

(e) 前記線スペクトル周波数とその長時間平均との差分に基づく第 1 の変動量を計算する処理と、

(f) 前記全帯域エネルギーとその長時間平均との差分に基づく第 2 の変動量

を計算する処理と、

(g) 前記低域エネルギーとその長時間平均との差分に基づく第 3 の変動量を計算する処理と、

(h) 前記零交叉数とその長時間平均との差分に基づく第 4 の変動量を計算する処理と、

(I) 前記第 1 の変動量の長時間平均を計算する処理と、

(j) 前記第 2 の変動量の長時間平均を計算する処理と、

(k) 前記第 3 の変動量の長時間平均を計算する処理と、

(l) 前記第 4 の変動量の長時間平均を計算する処理と、

の前記 (a) から (l) の処理を実行させるためのプログラムを記録した前記情報処理装置が読み取り可能な記録媒体。

【請求項 1 5】 請求項 1 3 又は請求項 1 4 に記載の記録媒体において、

(a) 過去に出力された前記判別の結果を保持する処理と、

(b) 前記第 1 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 5 のフィルタと第 6 のフィルタとを切り替える処理と、

(c) 前記第 2 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 7 のフィルタと第 8 のフィルタとを切り替える処理と、

(d) 前記第 3 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 9 のフィルタと第 1 0 のフィルタとを切り替える処理と、

(e) 前記第 4 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 1 1 のフィルタと第 1 2 のフィルタとを切り替える処理と、

の前記 (a) から (e) の処理を、前記情報処理装置に実行させるためのプログラムを記録した前記情報処理装置が読み取り可能な記録媒体。

【請求項 1 6】 請求項 1 3 から請求項 1 5 のいずれかに記載の記録媒体において、

前記特徴量として、過去に入力された前記音声信号から前記線スペクトル周波数と、前記全帯域エネルギーと、前記低域エネルギーと、前記零交叉数とを計算する処理を、前記情報処理装置に実行させるためのプログラムを記録した前記情報処理装置が読み取り可能な記録媒体。

【請求項 17】 請求項 13 から 16 のいずれかに記載の記録媒体において、

- (a) 前記音声信号から線スペクトル周波数 (LSF) を計算する処理と、
- (b) 前記音声信号から全帯域エネルギーを計算する処理と、
- (c) 前記音声信号から低域エネルギーを計算する処理と、
- (d) 前記音声信号から零交叉数を計算する処理と、

の前記(a)から(d)の処理のうちの少なくとも一つを、前記情報処理装置に実行させるためのプログラムを記録した前記情報処理装置が読み取り可能な記録媒体。

【請求項 18】 請求項 13 から 17 のいずれかに記載の記録媒体において、

- (a) 音声復号装置から過去に出力された再生音声信号を記憶保持する処理と、
- (b) 前記音声信号から線スペクトル周波数 (LSF) を計算する処理と、
- (c) 前記音声信号から全帯域エネルギーを計算する処理と、
- (d) 前記音声信号から低域エネルギーを計算する処理と、
- (e) 前記再生音声信号から零交叉数を計算する処理と、

の前記(a)の処理と、前記(b)から(e)の処理のうちの少なくとも一つを、前記情報処理装置に実行させるためのプログラムを記録した前記情報処理装置が読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声信号を低ビットレートで伝送するための符号化装置および復号装置において、符号化方法および復号方法を音声区間と非音声区間とで切り替える際に用いる音声検出方法および装置に関する。

【0002】

【従来の技術】

携帯電話などの移動体音声通信では会話音声の背景に雑音が存在するが、非音声区間における背景雑音を伝送するのに必要となるビットレートは音声に比べて低いと考えられる。このため、回線の使用効率向上の観点から、音声区間の検出を行い、非音声区間では背景雑音に特化したビットレートの低い符号化方式を使用することが多い。例えば、ITU-T 標準G.729音声符号化方式では、非音声区間では断続的に背景雑音についての少ない情報を伝送する。このとき、音声検出は、音声品質の劣化を回避し、かつビットレートを効果的に低減するために、正確に動作することが求められる。ここで、従来の音声検出方式として、例えば、「A Silence Compression Scheme for G.729 Optimized for Terminals Conforming to ITU-T V.70」(ITU-T Recommendation G.729, Annex B)（「文献1」という）、あるいは「ITU-T勧告V.70端末に適した標準JT-G729に対する無音圧縮手法」（電信電話技術委員会標準JT-G729、付属資料B）（「文献2」という）のB.3節（VADアルゴリズムの詳細記述）の記載、あるいは、「ITU-T Recommendation G.729 Annex B: A Silence Compression Scheme for Use with G.729 Optimized for V.70 Digital Simultaneous Voice and Data Applications」（IEEE Communication Magazine, pp.64-73, September 1997）（「文献3」という）が参照される。

【0003】

図6は、従来の音声検出装置の構成例を示すブロック図である。この音声検出装置への音声の入力は、 T_{fr} msec（例えば、10 msec）周期のブロック単位（フレーム）で行われるものとする。フレーム長を L_{fr} サンプル（例えば、80サンプル）とする。1フレームのサンプル数は、入力音声のサンプリング周波数（例えば、8kHz）によって定まる。

【0004】

図5を参照して、従来の音声検出装置の各構成要素について説明する。

【0005】

入力端子10から音声を入力し、入力端子11から線形予測係数を入力する。ここで、線形予測係数は、音声検出装置が用いられる音声符号化装置において、

前記入力音声ベクトルを線形予測分析して求められる。線形予測分析に関しては、周知の方法、例えば、L. R. Rabinerらによる「Digital Processing of Speech Signals」(Prentice-Hall, 1978) (「文献4」という)の第8章「Linear Predictive Coding of Speech」を参照できる。なお、本発明による音声検出装置が、音声符号化装置とは独立に実現される場合には、前記線形予測分析が該音声検出装置において実行される。

【0006】

LSF計算回路1011は、入力端子11を介して線形予測係数を入力し、前記線形予測係数から線スペクトル周波数(Line Spectral Frequency: LSF)を計算し、前記LSFを第1の変動量計算回路1031と第1の移動平均計算回路1021とへ出力する。ここで、線形予測係数からのLSFの計算に関しては、周知の方法、例えば、文献1の3.2.3節に記述されている方法等が用いられる。

【0007】

全帯域エネルギー計算回路1012は、入力端子10を介して音声(入力音声)を入力し、入力音声の全帯域エネルギーを計算し、前記全帯域エネルギーを第2の変動量計算回路1032と第2の移動平均計算回路1022とへ出力する。ここで、全帯域エネルギー E_f は、正規化された0次の自己相関関数 $R(0)$ の対数をとったものであり、次式で表される。

$$E_f = 10 \cdot \log_{10} \left[\frac{1}{N} R(0) \right]$$

また、自己相関係数は、次式で表される。

$$R(k) = \sum_{n=k}^{N-1} s^1(n) s^1(n-k)$$

ここで、 N は入力音声に対する線形予測分析の窓の長さ(分析窓長、例えば、240サンプル)であり、 $s^1(n)$ は、前記窓をかけた入力音声である。

【0008】

$N > L_{fr}$ の場合は、過去のフレームにおいて入力された音声を保持することに

より、前記分析窓長分の音声とする。

【 0 0 0 9 】

低域エネルギー計算回路 1 0 1 3 は、入力端子 1 0 を介して音声（入力音声）を入力し、入力音声の低域エネルギーを計算し、前記低域エネルギーを第 3 の変動量計算回路 1 0 3 3 と第 3 の移動平均計算回路 1 0 2 3 とへ出力する。ここで、0 から F_1 Hz までの低域エネルギー E_1 は、次式で表される。

$$E_1 = 10 \log_{10} \left[\frac{1}{N} \sum_{k=0}^{N-1} |x(k)|^2 \right]$$

ここで、

はカットオフ周波数が F_1 Hz の FIR フィルタのインパルス応答であり、

は対角成分が自己相関係数 $R(k)$ であるテプリッツ自己相関行列である。

【 0 0 1 0 】

零交叉数計算回路 1 0 1 4 は、入力端子 1 0 を介して音声（入力音声）を入力し、入力音声ベクトルの零交叉数を計算し、前記零交叉数を第 4 の変動量計算回路 1 0 3 4 と第 4 の移動平均計算回路 1 0 2 4 とへ出力する。ここで、零交叉数 Z_c は、次式で表される。

$$Z_c = \frac{1}{2\pi} \sum_{n=0}^{N-1} |\text{sgn}[x(n)] - \text{sgn}[x(n-1)]|$$

ここで、 $S(n)$ は入力音声であり、 $\text{sgn}[x]$ は x が正のとき 1 を、負のとき 0 をとる関数である。

【 0 0 1 1 】

第 1 の移動平均計算回路 1 0 2 1 は、LSF 計算回路 1 0 1 1 から LSF を入力し、前記 LSF と過去のフレームにおいて計算された平均 LSF とから現在のフレーム（現

フレーム)における平均LSFを計算し、これを第1の変動量計算回路1031へ出力する。ここで、第mフレームにおけるLSFを

$$L_{f,m}$$

とすると、第mフレームにおける平均LSF、

$$\bar{L}_m$$

は次式で表される。

$$\bar{L}_m = \frac{1}{P} \sum_{p=0}^{P-1} L_{f,m-p}$$

ここで、Pは線形予測次数(例えば、10)であり、 β_{LSF} はある定数(例えば、0.7)である。

【0012】

第2の移動平均計算回路1022は、全帯域エネルギー計算回路1012から全帯域エネルギーを入力し、前記全帯域エネルギーと過去のフレームにおいて計算された平均全帯域エネルギーとから現フレームにおける平均全帯域エネルギーを計算し、これを第2の変動量計算回路1032へ出力する。ここで、第mフレームにおける全帯域エネルギーを $E_f^{[m]}$ とすると、第mフレームにおける平均全帯域エネルギー

$$\bar{E}_m$$

は次式で表される。

$$\bar{E}_m = \beta_{Ef} \bar{E}_{m-1} + (1 - \beta_{Ef}) E_f^{[m]}$$

ここで、 β_{Ef} はある定数(例えば、0.7)である。

【0013】

第3の移動平均計算回路1023は、低域エネルギー計算回路1013から低域エネルギーを入力し、前記低域エネルギーと過去のフレームにおいて計算された平均低域エネルギーとから現フレームにおける平均低域エネルギーを計算し、これを第3の変動量計算回路1033へ出力する。ここで、第mフレームにお

る低域エネルギーを $E_1^{[m]}$ とすると、第 m フレームにおける平均低域エネルギー
は次式で表される。

$$E_1^{[m]} = \beta_{E1} \cdot E_1^{[m-1]} + (1 - \beta_{E1}) \cdot E_1^{[m]}$$

ここで、 β_{E1} はある定数（例えば、0.7）である。

【0 0 1 4】

第 4 の移動平均計算回路 1 0 2 4 は、零交叉数計算回路 1 0 1 4 から零交叉数
を入力し、前記零交叉数と過去のフレームにおいて計算された平均零交叉数と
から現フレームにおける平均零交叉数を計算し、これを第 4 の変動量計算回路 1 0
3 4 へ出力する。ここで、第 m フレームにおける零交叉数を $Z_c^{[m]}$ とすると、第
 m フレームにおける平均零交叉数

は次式で表される。

$$Z_c^{[m]} = \beta_{Zc} \cdot Z_c^{[m-1]} + (1 - \beta_{Zc}) \cdot Z_c^{[m]}$$

ここで、 β_{Zc} はある定数（例えば、0.7）である。

【0 0 1 5】

第 1 の変動量計算回路 1 0 3 1 は、LSF 計算回路 1 0 1 1 から LSF、 $\alpha_i^{[m]}$ を入
力し、第 1 の移動平均計算回路 1 0 2 1 から平均 LSF

を入力し、前記 LSF と前記平均 LSF とから、スペクトル変動量（第 1 の変動量）を
計算し、前記第 1 の変動量を音声/非音声判定回路 1 0 4 0 へ出力する。ここで
、第 m フレームにおける第 1 の変動量 $\Delta S^{[m]}$ は、次式で表される。

$$\Delta S^{[m]} = \sum_{i=1}^N (\alpha_i^{[m]} - \bar{\alpha}^{[m]})^2$$

第 2 の変動量計算回路 1 0 3 2 は、全帯域エネルギー計算回路 1 0 1 2 から全帯域エネルギー $E_f^{[m]}$ を入力し、第 2 の移動平均計算回路 1 0 2 2 から平均全帯域エネルギー

を

入力し、前記全帯域エネルギーと前記平均全帯域エネルギーとから全帯域エネルギー変動量（第 2 の変動量）を計算し、前記第 2 の変動量を音声/非音声判定回路 1 0 4 0 へ出力する。ここで、第 m フレームにおける第 2 の変動量 $\Delta E_f^{[m]}$ は、次式で表される。

は、

第 3 の変動量計算回路 1 0 3 3 は、低域エネルギー計算回路 1 0 1 3 から低域エネルギー $E_l^{[m]}$ を入力し、第 3 の移動平均計算回路 1 0 2 3 から平均低域エネルギー

を

入力し、前記低域エネルギーと前記平均低域エネルギーとから低域エネルギー変動量（第 3 の変動量）を計算し、前記第 3 の変動量を音声/非音声判定回路 1 0 4 0 へ出力する。ここで、第 m フレームにおける第 3 の変動量 $\Delta E_l^{[m]}$ は次式で表される。

は、

第 4 の変動量計算回路 1 0 3 4 は、零交叉数計算回路 1 0 1 4 から零交叉数 $Z_c^{[m]}$ を入力し、第 4 の移動平均計算回路 1 0 2 4 から平均零交叉数

を

入力し、前記零交叉数と前記平均零交叉数とから零交叉数変動量（第 4 の変動量）を計算し、前記第 4 の変動量を音声/非音声判定回路 1 0 4 0 へ出力する。ここで、第 m フレームにおける第 4 の変動量 $\Delta Z_c^{[m]}$ は次式で表される。

αH-2H-2H

音声/非音声判定回路 1 0 4 0 は、第 1 の変動量計算回路 1 0 3 1 から第 1 の変動量を入力し、第 2 の変動量計算回路 1 0 3 2 から第 2 の変動量を入力し、第 3 の変動量計算回路 1 0 3 3 から第 3 の変動量を入力し、第 4 の変動量計算回路 1 0 3 4 から第 4 の変動量を入力し、前記第 1 の変動量と、前記第 2 の変動量と、前記第 3 の変動量と、前記第 4 の変動量とからなる 4 次元ベクトルが、4 次元空間の音声領域内に存在するときは音声区間と判定し、それ以外の場合は非音声区間と判定し、前記音声区間のときは判定フラグを 1 に設定し、前記非音声区間のときは判定フラグを 0 に設定し、前記判定フラグを判定値平滑化回路 1 0 5 0 へ出力する。音声と非音声の判定（音声/非音声判定）には、例えば、文献 1 および 2 の B.3.5 節に記載されている 1 4 の境界判定を用いることができる。

【 0 0 1 6 】

判定値補正回路 1 0 5 0 は、音声/非音声判定回路 1 0 4 0 から判定フラグを入力し、全帯域エネルギー計算回路 1 0 1 2 から全帯域エネルギーを入力し、前記判定フラグをあらかじめ定められた条件式に従って補正し、補正された判定フラグを出力端子 1 2 を介して出力する。ここで、前記判定フラグの補正は以下のように行われる。前フレームが音声区間（すなわち判定フラグが 1）であり、かつ現フレームのエネルギーがある閾値を越えていれば、判定フラグを 1 とする。また、前フレームを含む 2 フレームが連続して音声区間であり、かつ現フレームのエネルギーと前フレームのエネルギーとの差分の絶対値がある閾値未満であれば、判定フラグを 1 とする。一方、過去の 1 0 フレームが非音声区間（すなわち判定フラグが 0）であり、かつ現フレームのエネルギーと前フレームのエネルギーとの差分がある閾値未満であれば、判定フラグを 0 とする。判定フラグの補正には、例えば、文献 1 および 2 の B.3.6 節に記載されている条件式を用いることができる。

【 0 0 1 7 】

【発明が解決しようとする課題】

上述した従来の音声検出方式は、音声区間における検出誤り（音声区間を誤って非音声区間と検出すること）および非音声区間における検出誤り（非音声区間を誤って音声区間と検出すること）を生じる場合がある、という問題点を有している。

【 0 0 1 8 】

その理由は、スペクトルの変動量、エネルギーの変動量および零交叉数の変動量を直接用いて音声/非音声判定を行うためである。実際の入力音声が発声区間であっても、前記各変動量の値は変動が大きいため、音声区間に対応するようにあらかじめ定めた値域に存在するとは限らない。よって、音声区間における前記検出誤りが生じる。このことは、非音声区間内においても同様である。

【 0 0 1 9 】

【課題を解決するための手段】

本願の第1の発明は、一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出方法において、前記特徴量の変動量を、フィルタに入力することで得られる前記変動量の長時間平均を用いることを特徴とする。

【 0 0 2 0 】

本願の第2の発明は、第1の発明において、前記特徴量の変動量を前記特徴量とその長時間平均とを用いて計算することを特徴とする。

【 0 0 2 1 】

本願の第3の発明は、第1または第2の発明において、前記音声検出方法によって過去に出力された前記判別の結果を用いて、前記変動量の長時間平均を計算する際の前記フィルタを切り替えることを特徴とする。

【 0 0 2 2 】

本願の第4の発明は、第1、第2または第3の発明において、過去に入力された前記音声信号から計算される特徴量を用いることを特徴とする。

【 0 0 2 3 】

本願の第5の発明は、第1、第2、第3または第4の発明において、前記特徴量として線スペクトル周波数、全帯域エネルギー、低域エネルギーおよび零交叉

数のうちの少なくとも一つを用いることを特徴とする。

本願の第6の発明は、第5の発明において、音声復号方法によって復号される線形予測係数から計算される線スペクトル周波数と、前記音声復号方法によって過去に出力された再生音声信号から計算される全帯域エネルギー、低域エネルギーおよび零交叉数のうちの少なくとも一つを用いることを特徴とする。

【 0 0 2 4 】

本願の第7の発明は、一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出装置において、前記音声信号から線スペクトル周波数（LSF）を計算するLSF計算回路と、前記音声信号から全帯域エネルギーを計算する全帯域エネルギー計算回路と、前記音声信号から低域エネルギーを計算する低域エネルギー計算回路と、前記音声信号から零交叉数を計算する零交叉数計算回路と、前記線スペクトル周波数の変動量（第1の変動量）を計算する線スペクトル周波数変動量計算部と、前記全帯域エネルギーの変動量（第2の変動量）を計算する全帯域エネルギー変動量計算部と、前記低域エネルギーの変動量（第3の変動量）を計算する低域エネルギー変動量計算部と、前記零交叉数の変動量（第4の変動量）を計算する零交叉数変動量計算部と、前記第1の変動量の長時間平均を計算する第1のフィルタと、前記第2の変動量の長時間平均を計算する第2のフィルタと、前記第3の変動量の長時間平均を計算する第3のフィルタと、前記第4の変動量の長時間平均を計算する第4のフィルタと、を含んで構成されることを特徴とする。

本願の第8の発明は、一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出装置において、前記音声信号から線スペクトル周波数（LSF）を計算するLSF計算回路と、前記音声信号から全帯域エネルギーを計算する全帯域エネルギー計算回路と、前記音声信号から低域エネルギーを計算する低域エネルギー計算回路と、前記音声信号から零交叉数を計算する零交叉数計算回路と、前記線スペクトル周波数とその長時間平均との差分に基づく第1の変動量を計算する第1の変動量計算回路と、前記全帯域エネルギーとその長時間平均との差分に基づく第2

の変動量を計算する第 2 の変動量計算回路と、前記低域エネルギーとその長時間平均との差分に基づく第 3 の変動量を計算する第 3 の変動量計算回路と、前記零交叉数とその長時間平均との差分に基づく第 4 の変動量を計算する第 4 の変動量計算回路と、前記第 1 の変動量の長時間平均を計算する第 1 のフィルタと、前記第 2 の変動量の長時間平均を計算する第 2 のフィルタと、前記第 3 の変動量の長時間平均を計算する第 3 のフィルタと、前記第 4 の変動量の長時間平均を計算する第 4 のフィルタと、を含んで構成されることを特徴とする。

【 0 0 2 5 】

本願の第 9 の発明は、第 7 または第 8 の発明において、前記音声検出装置から過去に出力された前記判別の結果を保持する第 1 の記憶回路と、前記第 1 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 5 のフィルタと第 6 のフィルタとを切り替える第 1 の切替器と、前記第 2 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 7 のフィルタと第 8 のフィルタとを切り替える第 2 の切替器と、前記第 3 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 9 のフィルタと第 1 0 のフィルタとを切り替える第 3 の切替器と、前記第 4 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 1 1 のフィルタと第 1 2 のフィルタとを切り替える第 4 の切替器と、を含んで構成されることを特徴とする。

【 0 0 2 6 】

本願の第 1 0 の発明は、第 7、第 8 または第 9 の発明において、過去に入力された前記音声信号から前記線スペクトル周波数と、前記全帯域エネルギーと、前記低域エネルギーと、前記零交叉数と、を計算することを特徴とする。

本願の第 1 1 の発明は、第 7 から第 1 0 の発明のいずれかにおいて、特徴量として、線スペクトル周波数、全帯域エネルギー、低域エネルギーおよび零交叉数のうちの少なくとも一つを用いることを特徴とする。

【 0 0 2 7 】

本願の第 1 2 の発明は、第 7 から第 1 0 の発明のいずれかにおいて、音声復号

装置から過去に出力された再生音声信号を記憶保持する第2の記憶回路を備え、前記第2の記憶回路から出力される前記再生音声信号から計算される全帯域エネルギー、低域エネルギーおよび零交叉数と、前記音声復号装置において復号される線形予測係数から計算される線スペクトル周波数と、のうちの少なくとも一つを用いることを特徴とする。

【 0 0 2 8 】

本願の第13の発明は、一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出方法を実行するプログラムを記録した記録媒体において、(a) 前記音声信号から線スペクトル周波数 (LSF) を計算する処理と、(b) 前記音声信号から全帯域エネルギーを計算する処理と、(c) 前記音声信号から低域エネルギーを計算する処理と、(d) 前記音声信号から零交叉数を計算する処理と、(e) 前記線スペクトル周波数の変動量 (第1の変動量) を計算する処理と、(f) 前記全帯域エネルギーの変動量 (第2の変動量) を計算する処理と、(g) 前記低域エネルギーの変動量 (第3の変動量) を計算する処理と、(h) 前記零交叉数の変動量 (第4の変動量) を計算する処理と、(I) 前記第1の変動量の長時間平均を計算する処理と、(j) 前記第2の変動量の長時間平均を計算する処理と、(k) 前記第3の変動量の長時間平均を計算する処理と、(l) 前記第4の変動量の長時間平均を計算する処理と、の前記 (a) から (l) の処理をコンピュータで実行させるプログラムを記録した記録媒体を提供する。

本願の第14の発明は、一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出方法を実行するプログラムを記録した記録媒体において、(a) 前記音声信号から線スペクトル周波数 (LSF) を計算する処理と、(b) 前記音声信号から全帯域エネルギーを計算する処理と、(c) 前記音声信号から低域エネルギーを計算する処理と、(d) 前記音声信号から零交叉数を計算する処理と、(e) 前記線スペクトル周波数とその長時間平均との差分に基づく第1の変動量を計算する処理と、(f) 前記全帯域エネルギーとその長時間平均との差分に基づく第2の変動量を計算する処理と、(g) 前記低域エネルギーとその長時間平均との差

分に基づく第3の変動量を計算する処理と、(h) 前記零交叉数とその長時間平均との差分に基づく第4の変動量を計算する処理と、(I) 前記第1の変動量の長時間平均を計算する処理と、(j) 前記第2の変動量の長時間平均を計算する処理と、(k) 前記第3の変動量の長時間平均を計算する処理と、(l) 前記第4の変動量の長時間平均を計算する処理と、の前記(a) から (l) の処理をコンピュータで実行させるプログラムを記録した記録媒体を提供する。

本願の第15の発明は、第13または第14の発明において、(a) 過去に出力された前記判別の結果を保持する処理と、(b) 前記第1の変動量の長時間平均を計算する際に、前記第1の記憶回路から入力した前記判別の結果を用いて、第5のフィルタと第6のフィルタとを切り替える処理と、(c) 前記第2の変動量の長時間平均を計算する際に、前記第1の記憶回路から入力した前記判別の結果を用いて、第7のフィルタと第8のフィルタとを切り替える処理と、(d) 前記第3の変動量の長時間平均を計算する際に、前記第1の記憶回路から入力した前記判別の結果を用いて、第9のフィルタと第10のフィルタとを切り替える処理と、(e) 前記第4の変動量の長時間平均を計算する際に、前記第1の記憶回路から入力した前記判別の結果を用いて、第11のフィルタと第12のフィルタとを切り替える処理と、の前記(a) から (e) の処理を、前記コンピュータに実行させるためのプログラムを記録した記録媒体を提供する。

【0029】

本願の第16の発明は、第13、第14または第15の発明において、過去に入力された前記音声信号から前記線スペクトル周波数と、前記全帯域エネルギーと、前記低域エネルギーと、前記零交叉数と、を計算する処理を、前記コンピュータに実行させるためのプログラムを記録した記録媒体を提供する。

【0030】

本願の第17の発明は、第13から第16のいずれかの発明において、
 (a) 前記音声信号から線スペクトル周波数 (LSF) を計算する処理と、
 (b) 前記音声信号から全帯域エネルギーを計算する処理と、
 (c) 前記音声信号から低域エネルギーを計算する処理と、
 (d) 前記音声信号から零交叉数を計算する処理と、

の前記(a)から(d)の処理のうちの少なくとも一つを、前記情報処理装置に実行させるためのプログラムを記録した前記情報処理装置が読み取り可能な記録媒体を提供する。

本願の第18の発明は、第13から第17のいずれかの発明において、

- (a) 音声復号装置から過去に出力された再生音声信号を記憶保持する処理と、
- (b) 前記音声信号から線スペクトル周波数 (LSF) を計算する処理と、
- (c) 前記音声信号から全帯域エネルギーを計算する処理と、
- (d) 前記音声信号から低域エネルギーを計算する処理と、
- (e) 前記再生音声信号から零交叉数を計算する処理と、

の前記(a)の処理と、前記(b)から(e)の処理のうちの少なくとも一つを、前記情報処理装置に実行させるためのプログラムを記録した前記情報処理装置が読み取り可能な記録媒体を提供する。

【0031】

本発明では、スペクトル変動量、エネルギー変動量および零交叉数変動量の長時間平均を用いて音声/非音声判定を行う。前記各変動量の長時間平均は、前記各変動量そのものに比べて、音声および非音声の各々の区間内における値の変動が小さいため、前記長時間平均の値は、音声区間および非音声区間に対応するようにあらかじめ定めた値域に高い割合で存在する。したがって、音声区間における検出誤りおよび非音声区間における検出誤りを低減できる。

【0032】

【発明の実施の形態】

次に、本発明の実施の形態について図面を参照して詳細に説明する。

【0033】

図1は、本発明の音声検出装置の第1の実施の形態の構成を示す図である。図1において、図6と同一または同等の要素には、同一の参照符号が付されている。図1において、入力端子10および11、出力端子12、LSF計算回路1011、全帯域エネルギー計算回路1012、低域エネルギー計算回路1013、零交叉数計算回路1014、第1の移動平均計算回路1021、第2の移動平均計算回路1022、第3の移動平均計算回路1023、第4の移動平均計算回路1

024、第1の変動量計算回路1031、第2の変動量計算回路1032、第3の変動量計算回路1033、第4の変動量計算回路1034および音声/非音声判定回路1040は、図5に示した要素と同じであるので、これらの要素の説明は省略し、以下では主に、図5に示した構成との相違点について説明する。

【0034】

図1を参照すると、本発明の第1の実施の形態においては、図5に示した構成に、第1のフィルタ2061、第2のフィルタ2062、第3のフィルタ2063および第4のフィルタ2064が付加されている。本発明の第1の実施の形態において、図5の構成と同様、音声の入力は、 T_{fr} msec（例えば、10 msec）周期のブロック単位（フレーム）で行われるものとする。フレーム長を L_{fr} サンプル（例えば、80サンプル）とする。1フレームのサンプル数は、入力音声のサンプリング周波数（例えば、8 kHz）によって定まる。

【0035】

第1のフィルタ2061は、第1の変動量計算回路1031から第1の変動量を入力し、前記第1の変動量の平均値、中央値あるいは最頻値など、前記第1の変動量の平均的な挙動を反映した値、第1の平均変動量を計算し、前記第1の平均変動量を音声/非音声判定回路1040へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。

【0036】

ここでは、次式の平滑フィルタを用いて、第 m フレームにおける第1の変動量 $\Delta S^{[m]}$ と第 $(m-1)$ フレームにおける第1の平均変動量

$$\Delta S^{[m-1]}$$

とから、第 m フレームにおける第1の平均変動量

$$\Delta S^{[m]}$$

を計算する。

$$\Delta \bar{S}^{[m]} = \gamma_S \cdot \Delta \bar{S}^{[m-1]} + (1 - \gamma_S) \cdot \Delta S^{[m]}$$

ここで、 γ_S は定数であり、例えば、 $\gamma_S = 0.74$ である。

【0037】

第2のフィルタ2062は、第2の変動量計算回路1032から第2の変動量を入力し、前記第2の変動量の平均値、中央値あるいは最頻値など、前記第2の変動量の平均的な挙動を反映した値、第2の平均変動量を計算し、前記第2の平均変動量を音声/非音声判定回路1040へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。

【0038】

ここでは、次式の平滑フィルタを用いて、第mフレームにおける第2の変動量 $\Delta E_f^{[m]}$ と第(m-1)フレームにおける第2の平均変動量

$$\Delta \bar{E}_f^{[m-1]}$$

とから、第mフレームにおける第2の平均変動量

$$\Delta \bar{E}_f^{[m]}$$

を計算する。

$$\Delta \bar{E}_f^{[m]} = \gamma_{Ef} \cdot \Delta \bar{E}_f^{[m-1]} + (1 - \gamma_{Ef}) \cdot \Delta E_f^{[m]}$$

ここで、 γ_{Ef} は定数であり、例えば、 $\gamma_{Ef} = 0.6$ である。

【0039】

第3のフィルタ2063は、第3の変動量計算回路1033から第3の変動量を入力し、前記第3の変動量の平均値、中央値あるいは最頻値など、前記第3の変動量の平均的な挙動を反映した値、第3の平均変動量を計算し、前記第3の平

均変動量を音声/非音声判定回路1040へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。

【0040】

ここでは、次式の平滑フィルタを用いて、第mフレームにおける第3の変動量 $\Delta E_1^{[m]}$ と第(m-1)フレームにおける第3の平均変動量

$$\Delta \bar{E}_1^{[m-1]}$$

とから、第mフレームにおける第3の平均変動量

$$\Delta \bar{E}_1^{[m]}$$

を計算する。

$$\Delta \bar{E}_1^{[m]} = \gamma_{E1} \cdot \Delta \bar{E}_1^{[m-1]} + (1 - \gamma_{E1}) \cdot \Delta E_1^{[m]}$$

ここで、 γ_{E1} は定数であり、例えば、 $\gamma_{E1} = 0.6$ である。

【0041】

第4のフィルタ2064は、第4の変動量計算回路1034から第4の変動量を入力し、前記第4の変動量の平均値、中央値あるいは最頻値など、前記第4の変動量の平均的な挙動を反映した値、第4の平均変動量を計算し、前記第4の平均変動量を音声/非音声判定回路1040へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。

【0042】

ここでは、次式の平滑フィルタを用いて、第mフレームにおける第4の変動量 $\Delta Z_c^{[m]}$ と第(m-1)フレームにおける第4の平均変動量

$$\Delta \bar{Z}_c^{[m-1]}$$

とから、第 m フレームにおける第 4 の平均変動量

$$\Delta \bar{z}_c^{[m]}$$

を計算する。

$$\Delta \bar{z}_c^{[m]} = \gamma_{zc} \cdot \Delta \bar{z}_c^{[m-1]} + (1 - \gamma_{zc}) \Delta z_c^{[m]}$$

ここで、 γ_{zc} は定数であり、例えば、 $\gamma_{zc} = 0.7$ である。

【 0 0 4 3 】

なお、第 1 の変動量計算回路 1 0 3 1、第 2 の変動量計算回路 1 0 3 2、第 3 の変動量計算回路 1 0 3 3 および第 4 の変動量計算回路 1 0 3 4 において計算される、第 1 の変動量、第 2 の変動量、第 3 の変動量および第 4 の変動量は、各々、従来例で示した式の代わりに、次式を用いて計算することもできる。これは、以下において記述される他の実施の形態に対しても同様である。

$$\Delta s^{[m]} = \sum_{i=1}^p \frac{|\omega_i^{[m]} - \bar{\omega}_i^{[m]}|}{\bar{\omega}_i^{[m]}}$$

$$\Delta \bar{s}_f^{[m]} = \frac{|\bar{s}_f^{[m]} - s_f^{[m]}|}{\bar{s}_f^{[m]}}$$

$$\Delta \bar{s}_f^{[m]} = \frac{|\bar{s}_f^{[m]} - s_f^{[m]}|}{\bar{s}_f^{[m]}}$$

$$\Delta \bar{z}_c^{[m]} = \frac{|\bar{z}_c^{[m]} - z_c^{[m]}|}{\bar{z}_c^{[m]}}$$

あるいは、次式を用いることもできる。

$$\Delta S_f^{[m]} = \sum_{i=1}^N \frac{(a_i^{[m]} - \bar{a}_f^{[m]})^2}{\bar{a}_f^{[m]}}$$

$$\Delta S_f^{[m]} = \frac{(\bar{S}_f^{[m]} - S_f^{[m]})^2}{\bar{S}_f^{[m]}}$$

$$\Delta S_f^{[m]} = \frac{(\bar{S}_f^{[m]} - S_f^{[m]})^2}{\bar{S}_f^{[m]}}$$

m

次に本発明の第2の実施の形態について説明する。図2は、本発明の音声検出装置の第2の実施の形態の構成を示す図である。図2において、図1および図6と同一または同等の要素には、同一の参照符号が付されている。

【0044】

図2を参照すると、本発明の第2の実施の形態では、第1の変動量と、第2の変動量と、第3の変動量と、第4の変動量の各々について平均値を計算するフィルタを、音声/非音声判定回路1040の出力に従って切り替える。ここで、平均値を計算するフィルタを前記第1の実施の形態と同様の平滑フィルタとすると、平滑化の強さを制御するパラメータ（平滑化強度パラメータ）、 γ_s 、 γ_{Ef} 、 γ_{El} およびmを、音声区間（すなわち、音声/非音声判定回路1040から出力される判定フラグが1）では大きくする。このことにより、前記第1の変動量および各差分の平均値が、音声区間の全体的な性質をよりよく反映することになり、音声区間での検出誤りをさらに低減できる。他方、非音声区間（すなわち、前

記判定フラグが0) では前記平滑化強度パラメータを小さくすることで、非音声区間から音声区間への遷移において、前記第1の変動量および各差分が平滑化されることで生じる判定フラグの遷移の遅れ、すなわち検出誤り、を回避できる。

【0045】

なお、入力端子10および11、出力端子12、LSF計算回路1011、全帯域エネルギー計算回路1012、低域エネルギー計算回路1013、零交叉数計算回路1014、第1の移動平均計算回路1021、第2の移動平均計算回路1022、第3の移動平均計算回路1023、第4の移動平均計算回路1024、第1の変動量計算回路1031、第2の変動量計算回路1032、第3の変動量計算回路1033、第4の変動量計算回路1034および音声/非音声判定回路1040は、図5に示した要素と同じであるので、これらの要素の説明は省略する。

【0046】

図2を参照すると、本発明の第2の実施の形態においては、図1に示した第1の実施の形態の構成における第1のフィルタ2061、第2のフィルタ2062、第3のフィルタ2063および第4のフィルタ2064に代わり、第5のフィルタ3061、第6のフィルタ3062、第7のフィルタ3063、第8のフィルタ3064、第9のフィルタ3065、第10のフィルタ3066、第11のフィルタ3067、第12のフィルタ3068、第1の切替器3071、第2の切替器3072、第3の切替器3073、第4の切替器3074、および第1の記憶回路3081、が付加されている。以下ではこれらについて説明する。

【0047】

第1の記憶回路3081は、音声/非音声判定回路1040から判定フラグを入力し、これを記憶保持し、記憶保持されている過去のフレームにおける前記判定フラグを第1の切替器3071と、第2の切替器3072と、第3の切替器3073と、第4の切替器3074とに出力する。

【0048】

第1の切替器3071は、第1の変動量計算回路1031から第1の変動量を入力し、第1の記憶回路3081から過去のフレームにおける判定フラグを入力

し、前記判定フラグが1（音声区間）のときは、前記第1の変動量を第5のフィルタ3061へ出力し、前記判定フラグが0（非音声区間）のときは、前記第1の変動量を第6のフィルタ3062へ出力する。

【0049】

第5のフィルタ3061は、第1の切替器3071から第1の変動量を入力し、前記第1の変動量の平均値、中央値あるいは最頻値など、前記第1の変動量の平均的な挙動を反映した値、第1の平均変動量を計算し、前記第1の平均変動量を音声/非音声判定回路1040へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。ここでは、次式の平滑化フィルタを用いて、第mフレームにおける第1の変動量 $\Delta S_{[m]}$ と第(m-1)フレームにおける第1の平均変動量

$$\Delta \bar{S}^{[m-1]}$$

とから、第mフレームにおける第1の平均変動量

$$\Delta \bar{S}^{[m]}$$

を計算する。

$$\Delta \bar{S}^{[m]} = r_{sl} \cdot \Delta \bar{S}^{[m-1]} + (1 - r_{sl}) \Delta S^{[m]}$$

ここで、 r_{sl} は定数であり、例えば、 $r_{sl} = 0.80$ である。

【0050】

第6のフィルタ3062は、第1の切替器3071から第1の変動量を入力し、前記第1の変動量の平均値、中央値あるいは最頻値など、前記第1の変動量の平均的な挙動を反映した値、第1の平均変動量を計算し、前記第1の平均変動量を音声/非音声判定回路1040へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができ

る。ここでは、次式の平滑化フィルタを用いて、第mフレームにおける第1の変動量 $\Delta S^{[m]}$ と第(m-1)フレームにおける第1の平均変動量

$$\Delta \bar{S}^{[m-1]}$$

とから、第mフレームにおける第1の平均変動量

$$\Delta \bar{S}^{[m]}$$

を計算する。

$$\Delta \bar{S}^{[m]} = \gamma_{s2} \cdot \Delta \bar{S}^{[m-1]} + (1 - \gamma_{s2}) \Delta S^{[m]}$$

ここで、 γ_{s2} は定数である。ただし、

$$\gamma_{s2} \leq \gamma_{s1}$$

例えば、 $\gamma_{s2} = 0.64$ である。

【0051】

第2の切替器3072は、第2の変動量計算回路1032から第2の変動量を入力し、第1の記憶回路3081から過去のフレームにおける判定フラグを入力し、前記判定フラグが1（音声区間）のときは、前記第2の変動量を第7のフィルタ3063へ出力し、前記判定フラグが0（非音声区間）のときは、前記第2の変動量を第8のフィルタ3064へ出力する。

【0052】

第7のフィルタ3063は、第2の切替器3072から第2の変動量を入力し、前記第2の変動量の平均値、中央値あるいは最頻値など、前記第2の変動量の平均的な挙動を反映した値、第2の平均変動量を計算し、前記第2の平均変動量を音声/非音声判定回路1040へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。ここでは、次式の平滑化フィルタを用いて、第mフレームにおける第2の変

動量 $\Delta E_f^{[m]}$ と第 $(m-1)$ フレームにおける第 2 の平均変動量

$$\Delta \bar{E}_f^{[m-1]}$$

とから、第 m フレームにおける第 2 の平均変動量

$$\Delta \bar{E}_f^{[m]}$$

を計算する。

$$\Delta \bar{E}_f^{[m]} = \gamma_{Ef1} \cdot \Delta \bar{E}_f^{[m-1]} + (1 - \gamma_{Ef1}) \cdot \Delta E_f^{[m]}$$

ここで、 γ_{Ef1} は定数であり、例えば、 $\gamma_{Ef1} = 0.70$ である。

【0053】

第 8 のフィルタ 3064 は、第 2 の切替器 3072 から第 2 の変動量を入力し、前記第 2 の変動量の平均値、中央値あるいは最頻値など、前記第 2 の変動量の平均的な挙動を反映した値、第 2 の平均変動量を計算し、前記第 2 の平均変動量を音声/非音声判定回路 1040 へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。ここでは、次式の平滑化フィルタを用いて、第 m フレームにおける第 2 の変動量 $\Delta E_f^{[m]}$ と第 $(m-1)$ フレームにおける第 2 の平均変動量

$$\Delta \bar{E}_f^{[m-1]}$$

とから、第 m フレームにおける第 2 の平均変動量

$$\Delta \bar{E}_f^{[m]}$$

を計算する。

$$\Delta \bar{E}_f^{[m]} = \gamma_{Ef2} \cdot \Delta \bar{E}_f^{[m-1]} + (1 - \gamma_{Ef2}) \cdot \Delta E_f^{[m]}$$

ここで、 γ_{Ef2} は定数であり、但し、

$$\gamma_{E2} \leq \gamma_{E1}$$

例えば、 $\gamma_{Ef2} = 0.54$ である。

【0054】

第3の切替器3073は、第3の変動量計算回路1033から第3の変動量を入力し、第1の記憶回路3081から過去のフレームにおける判定フラグを入力し、前記判定フラグが1（音声区間）のときは、前記第3の変動量を第9のフィルタ3065へ出力し、前記判定フラグが0（非音声区間）のときは、前記第3の変動量を第10のフィルタ3066へ出力する。

【0055】

第9のフィルタ3065は、第3の切替器3073から第3の変動量を入力し、前記第3の変動量の平均値、中央値あるいは最頻値など、前記第3の変動量の平均的な挙動を反映した値、第3の平均変動量を計算し、前記第3の平均変動量を音声/非音声判定回路1040へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。ここでは、次式の平滑化フィルタを用いて、第 m フレームにおける第3の変動量 $\Delta E_1^{[m]}$ と第 $(m-1)$ フレームにおける第3の平均変動量

$$\Delta \bar{E}_1^{[m-1]}$$

とから、第 m フレームにおける第3の平均変動量

$$\Delta \bar{E}_1^{[m]}$$

を計算する。

$$\Delta \bar{E}_1^{[m]} = \gamma_{Ef1} \cdot \Delta \bar{E}_1^{[m-1]} + (1 - \gamma_{Ef1}) \cdot \Delta E_1^{[m]}$$

ここで、 γ_{Ef1} は定数であり、例えば、 $\gamma_{Ef1} = 0.70$ である。

【 0 0 5 6 】

第 1 0 のフィルタ 3 0 6 6 は、第 3 の切替器 3 0 7 3 から第 3 の変動量を入力し、前記第 3 の変動量の平均値、中央値あるいは最頻値など、前記第 3 の変動量の平均的な挙動を反映した値、第 3 の平均変動量を計算し、前記第 3 の平均変動量を音声/非音声判定回路 1 0 4 0 へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。ここでは、次式の平滑化フィルタを用いて、第 m フレームにおける第 3 の変動量 $\Delta E_1^{[m]}$ と第 $(m-1)$ フレームにおける第 3 の平均変動量

$$\Delta \bar{E}_1^{[m-1]}$$

とから、第 m フレームにおける第 3 の平均変動量

$$\Delta \bar{E}_1^{[m]}$$

を計算する。

$$\Delta \bar{E}_1^{[m]} = \gamma_{Ef2} \cdot \Delta \bar{E}_1^{[m-1]} + (1 - \gamma_{Ef2}) \cdot \Delta E_1^{[m]}$$

ここで、 γ_{Ef2} は定数であり、但し、

$$\gamma_{S2} \leq \gamma_{S1}$$

例えば、 $\gamma_{Ef2} = 0.54$ である。

【 0 0 5 7 】

第 4 の切替器 3 0 7 4 は、第 4 の変動量計算回路 1 0 3 4 から第 4 の変動量を入力し、第 1 の記憶回路 3 0 8 1 から過去のフレームにおける判定フラグを入力し、前記判定フラグが 1 (音声区間) のときは、前記第 4 の変動量を第 1 1 のフィルタ 3 0 6 7 へ出力し、前記判定フラグが 0 (非音声区間) のときは、前記第 4 の変動量を第 1 2 のフィルタ 3 0 6 8 へ出力する。

【 0 0 5 8 】

第 1 1 のフィルタ 3 0 6 7 は、第 4 の切替器 3 0 7 4 から第 4 の変動量を入力し、前記第 4 の変動量の平均値、中央値あるいは最頻値など、前記第 4 の変動量の平均的な挙動を反映した値、第 4 の平均変動量を計算し、前記第 4 の平均変動量を音声/非音声判定回路 1 0 4 0 へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。ここでは、次式の平滑フィルタを用いて、第 m フレームにおける第 4 の変動量 $\Delta Z_c^{[m]}$ と第 $(m-1)$ フレームにおける第 4 の平均変動量

$$\Delta \bar{Z}_c^{[m-1]}$$

とから、第 m フレームにおける第 4 の平均変動量

$$\Delta \bar{Z}_c^{[m]}$$

を計算する。

$$\Delta \bar{Z}_c^{[m]} = r_{Zc1} \cdot \Delta \bar{Z}_c^{[m-1]} + (1 - r_{Zc1}) \cdot \Delta Z_c^{[m]}$$

ここで、 r_{Zc1} は定数であり、例えば、 $r_{Zc1} = 0.78$ である。

【 0 0 5 9 】

第 1 2 のフィルタ 3 0 6 8 は、第 4 の切替器 3 0 7 4 から第 4 の変動量を入力し、前記第 4 の変動量の平均値、中央値あるいは最頻値など、前記第 4 の変動量の平均的な挙動を反映した値、第 4 の平均変動量を計算し、前記第 4 の平均変動量を音声/非音声判定回路 1 0 4 0 へ出力する。ここで、前記平均値、中央値あるいは最頻値の計算には、線形フィルタおよび非線形フィルタを用いることができる。ここでは、次式の平滑フィルタを用いて、第 m フレームにおける第 4 の変動量 $\Delta Z_c^{[m]}$ と第 $(m-1)$ フレームにおける第 4 の平均変動量

$$\Delta \bar{Z}_c^{[m-1]}$$

とから、第 m フレームにおける第 4 の平均変動量

$$\Delta \bar{z}_c^{[m]}$$

を計算する。

$$\Delta \bar{z}_c^{[m]} = \gamma_{zc2} \cdot \Delta \bar{z}_c^{[m-1]} + (1 - \gamma_{zc2}) \cdot \Delta z_c^{[m]}$$

ここで、 γ_{zc2} は定数であり、例えば、

$$\gamma_{zc2} \leq \gamma_{zc1}$$

例えば、 $\gamma_{zc2} = 0.64$ である。

【0060】

次に本発明の第3の実施の形態について説明する。図3は、本発明の音声検出装置の第3の実施の形態の構成を示す図である。図3において、図1と同一または同等の要素には、同一の参照符号が付されている。本実施の形態は、例えば、音声復号装置において音声と非音声とに応じて復号処理方法を切り替える等の目的に対して、本願第1の実施の形態による音声検出装置を利用する場合の構成例、と位置付けられる。このために本実施の形態では、入力端子10を介して、前記音声復号装置から過去に出力された再生音声を入力し、入力端子11を介して、音声復号装置において復号された線形予測係数を入力する。なお、出力端子12、LSF計算回路1011、全帯域エネルギー計算回路1012、低域エネルギー計算回路1013、零交叉数計算回路1014、第1の移動平均計算回路1021、第2の移動平均計算回路1022、第3の移動平均計算回路1023、第4の移動平均計算回路1024、第1の変動量計算回路1031、第2の変動量計算回路1032、第3の変動量計算回路1033、第4の変動量計算回路1034、第1のフィルタ2061、第2のフィルタ2062、第3のフィルタ2063、第4のフィルタ2064および音声/非音声判定回路1040は、図1に示した要素と同じであるので、説明を省略する。

【0061】

図 3 を参照すると、本発明の第 3 の実施の形態は、図 1 に示した第 1 の実施の形態の構成に加えて、第 2 の記憶回路 7 0 7 1 を備えている。以下では、前記第 2 の記憶回路 7 0 7 1 について説明する。

【 0 0 6 2 】

第 2 の記憶回路 7 0 7 1 は、入力端子 1 0 を介して、音声復号装置から出力される再生音声を入力し、これを記憶保持し、記憶保持されている過去のフレームの再生信号を全帯域エネルギー計算回路 1 0 1 2 と、低域エネルギー計算回路 1 0 1 3 と、零交叉数計算回路 1 0 1 4 とへ出力する。

【 0 0 6 3 】

次に本発明の第 4 の実施の形態について説明する。図 4 は、本発明の音声検出装置の第 4 の実施の形態の構成を示す図である。図 4 において、図 2 と同一または同等の要素には、同一の参照符号が付されている。本実施の形態は、例えば、音声復号装置において音声と非音声とに応じて復号処理方法を切り替える等の目的に対して、本願第 2 の実施の形態による音声検出装置を利用する場合の構成例、と位置付けられる。このために本実施の形態では、入力端子 1 0 を介して、音声復号装置から出力される再生音声を入力し、入力端子 1 1 を介して、音声復号装置において復号された線形予測係数を入力する。なお、出力端子 1 2、LSF 計算回路 1 0 1 1、全帯域エネルギー計算回路 1 0 1 2、低域エネルギー計算回路 1 0 1 3、零交叉数計算回路 1 0 1 4、第 1 の移動平均計算回路 1 0 2 1、第 2 の移動平均計算回路 1 0 2 2、第 3 の移動平均計算回路 1 0 2 3、第 4 の移動平均計算回路 1 0 2 4、第 1 の変動量計算回路 1 0 3 1、第 2 の変動量計算回路 1 0 3 2、第 3 の変動量計算回路 1 0 3 3、第 4 の変動量計算回路 1 0 3 4、第 1 の切替器 3 0 7 1、第 2 の切替器 3 0 7 2、第 3 の切替器 3 0 7 3、第 4 の切替器 3 0 7 4、第 5 のフィルタ 3 0 6 1、第 6 のフィルタ 3 0 6 2、第 7 のフィルタ 3 0 6 3、第 8 のフィルタ 3 0 6 4、第 9 のフィルタ 3 0 6 5、第 1 0 のフィルタ 3 0 6 6、第 1 1 のフィルタ 3 0 6 7、第 1 2 のフィルタ 3 0 6 8、第 1 の記憶回路 3 0 8 1、および音声/非音声判定回路 1 0 4 0 は、図 2 に示した要素と同じであるので、説明を省略する。

【 0 0 6 4 】

図 4 を参照すると、本発明の第 4 の実施の形態は、図 2 に示した第 2 の実施の形態の構成に加えて、第 2 の記憶回路 7 0 7 1 を備えている。ここで、前記第 2 の記憶回路 7 0 7 1 は、図 3 に示した要素と同じであるので、説明を省略する。

【 0 0 6 5 】

上記した本発明の各実施の形態の音声検出装置は、ディジタル信号処理プロセッサ等のコンピュータ制御で実現するようにしてもよい。図 5 は、本発明の第 5 の実施の形態として、上記各実施の形態の音声検出装置をコンピュータで実現する場合の装置構成を模式的に示す図である。記録媒体 6 から読み出されたプログラムを実行するコンピュータ 1 において、一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出処理を実行するにあたり、記録媒体 6 には、

- (a) 前記音声信号から線スペクトル周波数 (LSF) を計算する処理と、
 - (b) 前記音声信号から全帯域エネルギーを計算する処理と、
 - (c) 前記音声信号から低域エネルギーを計算する処理と、
 - (d) 前記音声信号から零交叉数を計算する処理と、
 - (e) 前記線スペクトル周波数とその長時間平均との差分に基づく第 1 の変動量を計算する処理と、
 - (f) 前記全帯域エネルギーとその長時間平均との差分に基づく第 2 の変動量を計算する処理と、
 - (g) 前記低域エネルギーとその長時間平均との差分に基づく第 3 の変動量を計算する処理と、
 - (h) 前記零交叉数とその長時間平均との差分に基づく第 4 の変動量を計算する処理と、
 - (I) 前記第 1 の変動量の長時間平均を計算する処理と、
 - (j) 前記第 2 の変動量の長時間平均を計算する処理と、
 - (k) 前記第 3 の変動量の長時間平均を計算する処理と、
 - (l) 前記第 4 の変動量の長時間平均を計算する処理と、
- の前記 (a) から (l) の処理を実行させるためのプログラムが記録されている。

【 0 0 6 6 】

記録媒体 6 から該プログラムを記録媒体読出装置 5、記録媒体読出装置インタフェース 4 を介してメモリ 3 に読み出して実行する。上記プログラムは、マスク ROM 等、フラッシュ等の不揮発性メモリに格納してもよく、記録媒体は不揮発性メモリを含むほか、CD-ROM、FD、DVD (Digital Versatile Disk)、MT (磁気テープ)、可搬型 HDD 等の媒体のほか、例えばサーバ装置からコンピュータで該プログラムを通信媒体伝送する場合等、プログラムを担持する有線、無線で通信される通信媒体等も含む。

【 0 0 6 7 】

記録媒体 6 から読み出されたプログラムを実行するコンピュータ 1 において、一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出処理を実行するにあたり、記録媒体 6 には、

- (a) 過去に出力された前記判別の結果を保持する処理と、
 - (b) 前記第 1 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 5 のフィルタと第 6 のフィルタとを切り替える処理と、
 - (c) 前記第 2 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 7 のフィルタと第 8 のフィルタとを切り替える処理と、
 - (d) 前記第 3 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 9 のフィルタと第 10 のフィルタとを切り替える処理と、
 - (e) 前記第 4 の変動量の長時間平均を計算する際に、前記第 1 の記憶回路から入力した前記判別の結果を用いて、第 11 のフィルタと第 12 のフィルタとを切り替える処理と、
- の前記 (a) から (e) の処理を、前記コンピュータ 1 に実行させるためのプログラムが記録されている。

【 0 0 6 8 】

記録媒体 6 から読み出されたプログラムを実行するコンピュータ 1 において、

一定時間長毎に入力した音声信号から計算される特徴量を用いて、前記音声信号を一定時間長毎に音声区間と非音声区間とに判別する音声検出処理を実行するにあたり、記録媒体 6 には、過去に入力された前記音声信号から前記線スペクトル周波数と、前記全帯域エネルギーと、前記低域エネルギーと、前記零交叉数と、を計算する処理を、前記コンピュータ 1 に実行させるためのプログラムが記録されている。

【 0 0 6 9 】

記録媒体 6 から読み出されたプログラムを実行するコンピュータ 1 において、記録媒体 6 には、

- (a) 音声復号装置から過去に出力された再生音声信号を記憶保持する処理と、
 - (b) 前記再生音声信号から全帯域エネルギーを計算する処理と、
 - (c) 前記再生音声信号から低域エネルギーを計算する処理と、
 - (d) 前記再生音声信号から零交叉数を計算する処理と、
 - (e) 前記音声復号装置において復号される線形予測係数から線スペクトル周波数を計算する処理と、
- の前記 (a) から (e) の処理を、前記コンピュータに実行させるためのプログラムが記録されている。

【 0 0 7 0 】

次に、上述した処理の動作をフローチャートを用いて説明する。まず、上述した第 1 の実施の形態に相当する動作を説明する。図 7 は第 1 の実施の形態に相当する動作を説明する為のフローチャートである。

【 0 0 7 1 】

線形予測係数を入力し (Step 11)、前記線形予測係数から線スペクトル周波数 (Line Spectral Frequency: LSF) を計算する (Step A1)。ここで、線形予測係数からの LSF の計算に関しては、周知の方法、例えば、文献 1 の 3.2.3 節に記述されている方法等が用いられる。

【 0 0 7 2 】

次に、計算した LSF と過去のフレームにおいて計算された平均 LSF とから現在の

フレーム（現フレーム）における移動平均LSFを計算する（Step A2）。

【0073】

ここで、第mフレームにおけるLSFを

$$\alpha_i^{[m]}, i=1, \dots, P$$

とすると、第mフレームにおける平均LSF、

$$\bar{\alpha}_i^{[m]}, i=1, \dots, P$$

は次式で表される。

$$\bar{\alpha}_i^{[m]} = \beta_{LSF} \bar{\alpha}_i^{[m-1]} + (1 - \beta_{LSF}) \alpha_i^{[m]}, i=1, \dots, P$$

ここで、Pは線形予測次数（例えば、10）であり、 β_{LSF} はある定数（例えば、0.7）である。

【0074】

続いて、計算されたLSF $\alpha_i^{[m]}$ と移動平均LSF

$$\bar{\alpha}_i^{[m]}$$

とに基づいて、スペクトル変動量（第1の変動量）を計算する（Step A3）。

【0075】

ここで、第mフレームにおける第1の変動量 $\Delta S^{[m]}$ は次式で表される。

$$\Delta S^{[m]} = \sum_{i=1}^P \left(\alpha_i^{[m]} - \bar{\alpha}_i^{[m]} \right)^2$$

さらに、第1の変動量 $\Delta S^{[m]}$ から、前記第1の変動量の平均値、中央値あるいは最頻値など、前記第1の変動量の平均的な挙動を反映した値、第1の平均変動量を計算する（Step A3）。

【0076】

ここでは、次式の平滑フィルタを用いて、第 m フレームにおける第1の変動量 $\Delta S^{[m]}$ と第 $(m-1)$ フレームにおける第1の平均変動量

$$\Delta \bar{S}^{[m-1]}$$

とから、第 m フレームにおける第1の平均変動量

$$\Delta \bar{S}^{[m]}$$

を計算するものとする。

$$\Delta \bar{S}^{[m]} = \gamma_S \cdot \Delta \bar{S}^{[m-1]} + (1 - \gamma_S) \cdot \Delta S^{[m]}$$

ここで、 γ_S は定数であり、例えば、 $\gamma_S = 0.74$ である。

また、音声（入力音声）を入力し（Step 12）、入力音声の全帯域エネルギーを計算する（Step B1）。

【0077】

ここで、全帯域エネルギー E_f は、正規化された0次の自己相関関数 $R(0)$ の対数をとったものであり、次式で表される。

$$E_f = 10 \cdot \log_{10} \left[\frac{1}{N} R(0) \right]$$

また、自己相関係数は、次式で表される。

$$R(k) = \sum_{n=k}^{N-1} s^1(n) s^1(n-k)$$

ここで、 N は入力音声に対する線形予測分析の窓の長さ（分析窓長、例えば、

240サンプル) であり、 $S^1(n)$ は、前記窓をかけた入力音声である。 $N > L_{fr}$ の場合は、過去のフレームにおいて入力された音声を保持することにより、前記分析窓長分の音声とする。

【 0 0 7 8 】

次に、全帯域エネルギー E_f と過去のフレームにおいて計算された平均全帯域エネルギーとから現フレームにおける全帯域エネルギーの移動平均を計算する (Step B2)。

【 0 0 7 9 】

ここで、第 m フレームにおける全帯域エネルギーを $E_f^{[m]}$ とすると、第 m フレームにおける全帯域エネルギーの移動平均

$$\bar{E}_f^{[m]}$$

は次式で表される。

$$\bar{E}_f^{[m]} = \beta_{Ef} \cdot \bar{E}_f^{[m-1]} + (1 - \beta_{Ef}) \cdot E_f^{[m]}$$

ここで、 β_{Ef} はある定数 (例えば、0.7) である。

【 0 0 8 0 】

次に、全帯域エネルギー、 $E_f^{[m]}$ と、全帯域エネルギーの移動平均

$$\bar{E}_f^{[m]}$$

とから、全帯域エネルギー変動量 (第2の変動量) を計算する (Step B3)。

【 0 0 8 1 】

ここで、第 m フレームにおける第2の変動量 $\Delta E_f^{[m]}$ は、次式で表される。

$$\Delta E_f^{[m]} = \bar{E}_f^{[m]} - E_f^{[m]}$$

さらに、第 2 の変動量 $\Delta E_f^{[m]}$ から、第 2 の変動量の平均値、中央値あるいは最頻値など、前記第 2 の変動量の平均的な挙動を反映した値、第 2 の平均変動量を計算する (Step B4)。

【0082】

ここでは、次式の平滑フィルタを用いて、第 m フレームにおける第 2 の変動量 $\Delta E_f^{[m]}$ と第 $(m-1)$ フレームにおける第 2 の平均変動量

$$\Delta \bar{E}_f^{[m-1]}$$

とから、第 m フレームにおける第 2 の平均変動量

$$\Delta \bar{E}_f^{[m]}$$

を計算する。

$$\Delta \bar{E}_f^{[m]} = \frac{1}{l} \sum_{i=1}^l \Delta E_{f,i}^{[m]}$$

ここで、 γ_{Ef} は定数であり、例えば、 $\gamma_{Ef} = 0.6$ である。

【0083】

また、入力音声から、入力音声の低域エネルギーを計算する (Step C1) 。ここで、0 から F_1 Hz までの低域エネルギー E_l は、次式で表される。

$$E_l = 10 \cdot \log_{10} \left[\frac{1}{N} \sum_{n=0}^{N-1} \hat{h}^T R \hat{h} \right]$$

ここで、

$$\hat{h}$$

はカットオフ周波数が F_1 Hz の FIR フィルタのインパルス応答であり、

\hat{R}

は対角成分が自己相関係数 $R(k)$ であるテプリッツ自己相関行列である。

【 0 0 8 4 】

次に、低域エネルギーと過去のフレームにおいて計算された平均低域エネルギーとから現フレームにおける低域エネルギーの移動平均を計算する (S t e p C 2) 。ここで、第 m フレームにおける低域エネルギーを $E_1^{[m]}$ とすると、第 m フレームにおける平均低域エネルギー

 $\bar{E}_1^{[m]}$

は次式で表される。

$$\bar{E}_1^{[m]} = \beta_{E1} \bar{E}_1^{[m-1]} + (1 - \beta_{E1}) E_1^{[m]}$$

ここで、 β_{E1} はある定数 (例えば、0.7) である。

【 0 0 8 5 】

続いて、低域エネルギー $E_1^{[m]}$ と、低域エネルギーの移動平均

 $\bar{E}_1^{[m]}$

とから、低域エネルギー変動量 (第3の変動量) を計算する (S t e p C 3) 。ここで、第 m フレームにおける第3の変動量 $\Delta E_1^{[m]}$ は次式で表される。

$$\Delta E_1^{[m]} = \bar{E}_1^{[m]} - E_1^{[m]}$$

さらに、前記第3の変動量の平均値、中央値あるいは最頻値など、前記第3の変動量の平均的な挙動を反映した値、第3の平均変動量を計算する (S t e p C 4) 。ここでは、次式の平滑フィルタを用いて、第 m フレームにおける第3の

変動量 $\Delta E_1^{[m]}$ と第 $(m-1)$ フレームにおける第 3 の平均変動量

$$\Delta \bar{E}_1^{[m-1]}$$

とから、第 m フレームにおける第 3 の平均変動量

$$\Delta \bar{E}_1^{[m]}$$

を計算する。

$$\Delta \bar{E}_1^{[m]} = \Delta \bar{E}_1^{[m-1]} + \gamma_{E1} \Delta E_1^{[m]}$$

ここで、 γ_{E1} は定数であり、例えば、 $\gamma_{E1} = 0.6$ である。

【0086】

また、音声（入力音声）から入力音声ベクトルの零交叉数を計算する（Step D1）。ここで、零交叉数 Z_c は、次式で表される。

$$Z_c = \frac{1}{2L_f} \sum_{n=0}^{L_f-1} | \text{sgn}[s(n)] - \text{sgn}[s(n-1)] |$$

ここで、 $S(n)$ は入力音声であり、 $\text{sgn}[x]$ は x が正のとき 1 を、負のとき 0 をとる関数である。

【0087】

次に、計算した零交叉数と過去のフレームにおいて計算された平均零交叉数とから現フレームにおける零交叉数の移動平均を計算する（Step D2）。ここで、第 m フレームにおける零交叉数を

$$Z_c^{[m]}$$

とすると、第 m フレームにおける平均零交叉数

$$\Delta Z_c^{[m]}$$

は次式で表される。

$$\bar{z}_c^{[m]} = \beta_{zc} \cdot \bar{z}_c^{[m-1]} + \{1 - \beta_{zc}\} z_c^{[m]}$$

ここで、 β_{zc} はある定数（例えば、0.7）である。

【0088】

次に、零交叉数 $Z_c^{[m]}$ と、零交叉数の移動平均

$$\bar{z}_c^{[m]}$$

とをから、零交叉数変動量（第4の変動量）を計算する（Step D3）。ここで、第mフレームにおける第4の変動量 $\Delta Z_c^{[m]}$ は、次式で表される。

$$\Delta Z_c^{[m]} = \bar{z}_c^{[m]} - z_c^{[m]}$$

さらに、第4の変動量から、前記第4の変動量の平均値、中央値あるいは最頻値など、前記第4の変動量の平均的な挙動を反映した値、第4の平均変動量を計算する（Step D4）。ここでは、次式の平滑フィルタを用いて、第mフレームにおける第4の変動量 $\Delta Z_c^{[m]}$ と第 $(m-1)$ フレームにおける第4の平均変動量

$$\bar{\Delta Z}_c^{[m-1]}$$

とから、第mフレームにおける第4の平均変動量

$$\bar{\Delta Z}_c^{[m]}$$

を計算する。

$$\Delta \bar{z}_c^{[m]} = r_{zc} \cdot \Delta \bar{z}_c^{[m-1]} + (1 - r_{zc}) \Delta z_c^{[m]}$$

ここで、 r_{zc} は定数であり、例えば、 $r_{zc} = 0.7$ である。

【0089】

最後に、前記第1の平均変動量

$$\Delta \bar{z}_1^{[m]}$$

と、前記第2の平均変動量

$$\Delta \bar{z}_2^{[m]}$$

と、前記第3の平均変動量

$$\Delta \bar{z}_3^{[m]}$$

と、前記第4の平均変動量

$$\Delta \bar{z}_4^{[m]}$$

とからなる4次元ベクトルが、4次元空間の音声領域内に存在するときは音声区間と判定し、それ以外の場合は非音声区間と判定する（Step E1）。

【0090】

そして、前記音声区間のときは判定フラグを1に設定し（Step E3）、前記非音声区間のときは判定フラグを0に設定する（Step E2）し、判定結果を出力する（Step E4）。

【0091】

以上、処理が終了する。

【0092】

次に、上述した第2の実施の形態に相当する処理の動作をフローチャートを用いて説明する。図8、図9及び図10は第2の実施の形態に相当する動作を説明

する為のフローチャートである。尚、上述した動作と同じ処理については説明を省略し、異なるものについてのみ説明する。

【 0 0 9 3 】

上述した処理と異なるのは、第 1 の変動量、第 2 の変動量、第 3 の変動量及び第 4 の変動量を計算した後、これらの平均値を計算する際、判定フラグの種類により平均値を計算するフィルタを切り替える点である。

【 0 0 9 4 】

まず、第 1 の変動量の場合について説明する。

【 0 0 9 5 】

Step A3 で第 1 の変動量を計算した後、過去の判定フラグが 1 か否かを確認する (Step A11)。

【 0 0 9 6 】

判定フラグが 1 であれば、第 2 の実施の形態における第 5 のフィルタのようなフィルタ処理を行い、第 1 の平均変動量を計算する (Step A12)。例えば、次式の平滑化フィルタを用いて、第 m フレームにおける第 1 の変動量 $\Delta S^{[m]}$ と第 (m-1) フレームにおける第 1 の平均変動量

$$\Delta \bar{S}^{[m-1]}$$

とから、第 m フレームにおける第 1 の平均変動量

$$\Delta \bar{S}^{[m]}$$

を計算する。

$$\Delta \bar{S}^{[m]} = r_{s1} \cdot \Delta \bar{S}^{[m-1]} + (1 - r_{s1}) \Delta S^{[m]}$$

ここで、 r_{s1} は定数であり、例えば、 $r_{s1} = 0.80$ である。

【 0 0 9 7 】

一方、判定フラグが 0 であれば、第 2 の実施の形態における第 6 のフィルタの

ようなフィルタ処理を行い、第1の平均変動量を計算する (Step A13)。
例えば、次式の平滑化フィルタを用いて、第mフレームにおける第1の変動量 $\Delta S^{[m]}$ と第 (m-1) フレームにおける第1の平均変動量

$$\Delta S^{[m-1]}$$

とから、第mフレームにおける第1の平均変動量

$$\Delta S^{[m]}$$

を計算する。

$$\Delta S^{[m]} = r_{S2} \cdot \Delta S^{[m-1]} + \{1 - r_{S2}\} \Delta S^{[m]}$$

ここで、 r_{S2} は定数である。ただし、

$$r_{S2} \leq r_{S1}$$

例えば、 $r_{S2} = 0.64$ である。

【0098】

次に、第2の変動量の場合について説明する。

【0099】

Step B3で第2の変動量を計算した後、過去の判定フラグが1か否かを
確認する (Step B11)。

【0100】

判定フラグが1であれば、第2の実施の形態における第7のフィルタのような
フィルタ処理を行い、第2の平均変動量を計算する (Step B12)。例え
ば、次式の平滑化フィルタを用いて、第mフレームにおける第2の変動量 $\Delta E_f^{[m]}$
と第 (m-1) フレームにおける第2の平均変動量

$$\Delta E_f^{[m-1]}$$

とから、第mフレームにおける第2の平均変動量

$$\Delta \bar{E}_f^{[m]}$$

を計算する。

$$\Delta \bar{E}_f^{[m]} = \gamma_{Ef1} \cdot \Delta \bar{E}_f^{[m-1]} + (1 - \gamma_{Ef1}) \cdot \Delta E_f^{[m]}$$

ここで、 γ_{Ef1} は定数であり、例えば、 $\gamma_{Ef1} = 0.70$ である。

【0101】

一方、判定フラグが0であれば、第2の実施の形態における第8のフィルタのようなフィルタ処理を行い、第2の平均変動量を計算する（Step B13）。例えば、次式の平滑化フィルタを用いて、第mフレームにおける第2の変動量 $\Delta E_f^{[m]}$ と第(m-1)フレームにおける第2の平均変動量

$$\Delta \bar{E}_f^{[m-1]}$$

とから、第mフレームにおける第2の平均変動量

$$\Delta \bar{E}_f^{[m]}$$

を計算する。

$$\Delta \bar{E}_f^{[m]} = \gamma_{Ef2} \cdot \Delta \bar{E}_f^{[m-1]} + (1 - \gamma_{Ef2}) \cdot \Delta E_f^{[m]}$$

ここで、 γ_{Ef2} は定数であり、但し、

$$\gamma_{Ef2} \leq \gamma_{Ef1}$$

例えば、 $\gamma_{Ef2} = 0.54$ である。

【0102】

続いて、第 3 の変動量の場合について説明する。

【0103】

Step C3 で第 3 の変動量を計算した後、過去の判定フラグが 1 か否かを
確認する (Step C11)。

【0104】

判定フラグが 1 であれば、第 2 の実施の形態における第 9 のフィルタのような
フィルタ処理を行い、第 3 の平均変動量を計算する (Step C12)。例え
ば、次式の平滑化フィルタを用いて、第 m フレームにおける第 3 の変動量 $\Delta E_1^{[m]}$
と第 $(m-1)$ フレームにおける第 3 の平均変動量

$$\Delta \bar{E}_1^{[m-1]}$$

とから、第 m フレームにおける第 3 の平均変動量

$$\Delta \bar{E}_1^{[m]}$$

を計算する。

$$\Delta \bar{E}_1^{[m]} = \gamma_{Ef1} \cdot \Delta \bar{E}_1^{[m-1]} + (1 - \gamma_{Ef1}) \cdot \Delta E_1^{[m]}$$

ここで、 γ_{Ef1} は定数であり、例えば、 $\gamma_{Ef1} = 0.70$ である。

【0105】

一方、判定フラグが 0 であれば、第 2 の実施の形態における第 10 のフィルタ
のようなフィルタ処理を行い、第 3 の平均変動量を計算する (Step C13
)。例えば、次式の平滑化フィルタを用いて、第 m フレームにおける第 3 の変動
量 $\Delta E_1^{[m]}$ と第 $(m-1)$ フレームにおける第 3 の平均変動量

$$\Delta \bar{E}_1^{[m-1]}$$

とから、第 m フレームにおける第 3 の平均変動量

$$\Delta \bar{E}_1^{[m]}$$

を計算する。

$$\Delta \vec{B}_f^{[m]} = \gamma_{Ef2} \cdot \Delta \vec{B}_f^{[m-1]} + (1 - \gamma_{Ef2}) \cdot \Delta \vec{B}_f^{[m]}$$

ここで、 γ_{Ef2} は定数であり、

$$\gamma_{Ef2} \leq \gamma_{Ef1}$$

$\gamma_{Ef2} = 0.54$ である。

【0106】

さらに、第4の変動量の場合について説明する。

【0107】

Step D3で第4の変動量を計算した後、過去の判定フラグが1か否かを
確認する (Step D11)。

【0108】

判定フラグが1であれば、第2の実施の形態における第11のフィルタのよう
なフィルタ処理を行い、第4の平均変動量を計算する (Step D12)。例
えば、次式の平滑フィルタを用いて、第mフレームにおける第4の変動量 $\Delta Z_c^{[m]}$
と第(m-1)フレームにおける第4の平均変動量

$$\Delta \vec{Z}_c^{[m-1]}$$

とから、第mフレームにおける第4の平均変動量

$$\Delta \vec{Z}_c^{[m]}$$

を計算する。

$$\Delta \vec{Z}_c^{[m]} = \gamma_{Zc1} \cdot \Delta \vec{Z}_c^{[m-1]} + (1 - \gamma_{Zc1}) \cdot \Delta \vec{Z}_c^{[m]}$$

ここで、 γ_{zc1} は定数であり、例えば、 $\gamma_{zc1} = 0.78$ である。

【0109】

一方、判定フラグが0であれば、第2の実施の形態における第12のフィルタのようなフィルタ処理を行い、第4の平均変動量を計算する（Step D13）。例えば、次式の平滑フィルタを用いて、第mフレームにおける第4の変動量 $\Delta Z_c^{[m]}$ と第(m-1)フレームにおける第4の平均変動量

$$\Delta \bar{Z}_c^{[m-1]}$$

とから、第mフレームにおける第4の平均変動量

$$\Delta \bar{Z}_f^{[m]}$$

を計算する。

$$\Delta \bar{Z}_c^{[m]} = \gamma_{zc2} \cdot \Delta \bar{Z}_c^{[m-1]} + (1 - \gamma_{zc2}) \Delta Z_c^{[m]}$$

ここで、 γ_{zc2} は定数であり、但し、

$$\gamma_{zc2} \leq \gamma_{zc1}$$

$\gamma_{zc2} = 0.64$ である。

【0110】

そして、前記第1の平均変動量

$$\Delta \bar{S}^{[m]}$$

と、前記第2の平均変動量

$$\Delta \bar{E}_f^{[m]}$$

と、前記第3の平均変動量

$$\Delta \vec{b}^{[m]}$$

と、前記第 4 の平均変動量

$$\Delta \vec{b}^{[m]}$$

とからなる 4 次元ベクトルが、4 次元空間の音声領域内に存在するときは音声区
間と判定し、それ以外のときは非音声区間と判定する (Step E1)。

【0111】

続いて、上述した第 3 の実施の形態に相当する処理の動作をフローチャートを用いて説明する。図 11 は第 3 の実施の形態に相当する動作を説明する為のフローチャートである。

【0112】

本動作において上述した処理と異なるのは、Step I11 と Step I12 であり、Step I11 において音声復号装置において復号された線形予測係数を入力する点と、Step I12 において音声復号装置から過去に出力された再生音声ベクトルを入力する点とが異なる。

【0113】

これら以外は、上述した動作と同じ処理なので、説明を省略する。

【0114】

最後に、上述した第 4 の実施の形態に相当する処理の動作をフローチャートを用いて説明する。図 12、図 13 及び図 14 は第 4 の実施の形態に相当する動作を説明する為のフローチャートである。

【0115】

本動作は、上述した第 2 の実施の形態に相当する動作と第 3 の実施の形態に相当する動作を組み合わせたことを特徴とする。従って、第 2 の実施の形態に相当する動作と第 3 の実施の形態に相当する動作とは、既に説明しているので、詳細な説明は省略する。

【0116】

【発明の効果】

本発明の効果は、音声区間における検出誤りおよび非音声区間における検出誤りを低減できることである。

【0117】

その理由は、スペクトル変動量、エネルギー変動量および零交叉数変動量の長時間平均を用いて音声/非音声判定を行うからである。すなわち、前記各変動量そのものに比べて、前記各変動量の長時間平均は、音声および非音声の各々の区間内における値の変動が小さいため、前記長時間平均の値は、音声区間および非音声区間に対応するようにあらかじめ定めた値域に高い割合で存在するからである。

【図面の簡単な説明】

【図1】

本発明の音声検出装置の第1の実施の形態を示すブロック図である。

【図2】

本発明の音声検出装置の第2の実施の形態を示すブロック図である。

【図3】

本発明の音声検出装置の第3の実施の形態を示すブロック図である

【図4】

本発明の音声検出装置の第4の実施の形態を示すブロック図である

【図5】

本発明の第5の実施の形態を示すブロック図である。

【図6】

従来の音声検出装置を説明するブロック図である。

【図7】

本発明の実施の形態の動作を説明するためのフローチャートである。

【図8】

本発明の実施の形態の動作を説明するためのフローチャートである。

【図9】

本発明の実施の形態の動作を説明するためのフローチャートである。

【図10】

本発明の実施の形態の動作を説明するためのフローチャートである。

【図 1 1】

本発明の実施の形態の動作を説明するためのフローチャートである。

【図 1 2】

本発明の実施の形態の動作を説明するためのフローチャートである。

【図 1 3】

本発明の実施の形態の動作を説明するためのフローチャートである。

【図 1 4】

本発明の実施の形態の動作を説明するためのフローチャートである。

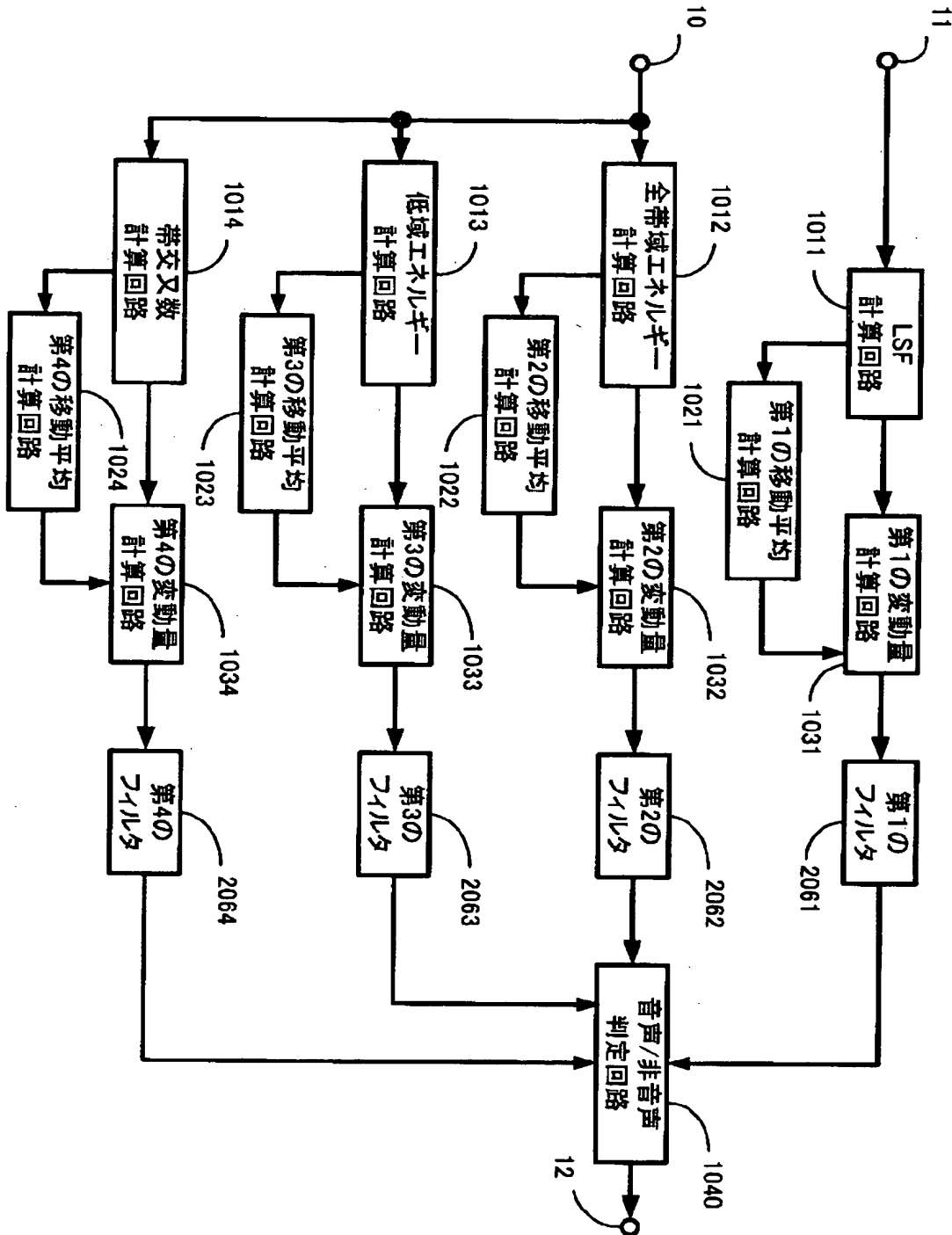
【符号の説明】

- 1 コンピュータ
- 2 CPU
- 3 メモリ
- 4 記録媒体読出装置インタフェース
- 5 記録媒体読出装置
- 6 記録媒体
- 10,11 入力端子
- 20 出力端子
- 1011 LSF計算回路
- 1012 全帯域エネルギー計算回路
- 1013 低域エネルギー計算回路
- 1014 零交叉数計算回路
- 1021 第 1 の移動平均計算回路
- 1022 第 2 の移動平均計算回路
- 1023 第 3 の移動平均計算回路
- 1024 第 4 の移動平均計算回路
- 1031 第 1 の変動量計算回路
- 1032 第 2 の変動量計算回路
- 1033 第 3 の変動量計算回路

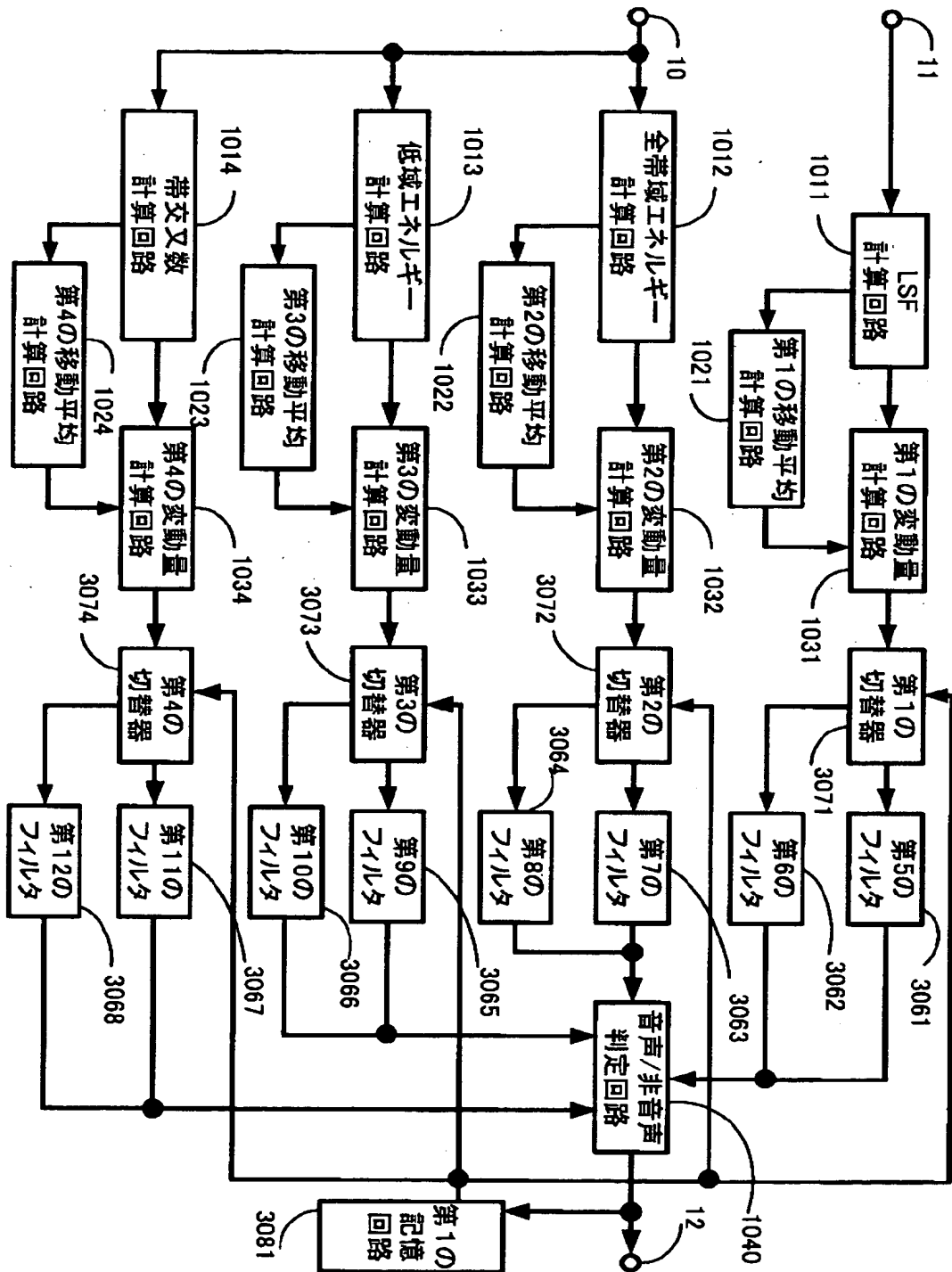
- 1034 第 4 の変動量計算回路
- 1040 音声/非音声判定回路
- 1050 判定値補正回路
- 2061 第 1 のフィルタ
- 2062 第 2 のフィルタ
- 2063 第 3 のフィルタ
- 2064 第 4 のフィルタ
- 3061 第 5 のフィルタ
- 3062 第 6 のフィルタ
- 3063 第 7 のフィルタ
- 3064 第 8 のフィルタ
- 3065 第 9 のフィルタ
- 3066 第 1 0 のフィルタ
- 3067 第 1 1 のフィルタ
- 3068 第 1 2 のフィルタ
- 3071 第 1 の切替器
- 3072 第 2 の切替器
- 3073 第 3 の切替器
- 3074 第 4 の切替器
- 3081 第 1 の記憶回路
- 7071 第 2 の記憶回路

【書類名】 図面

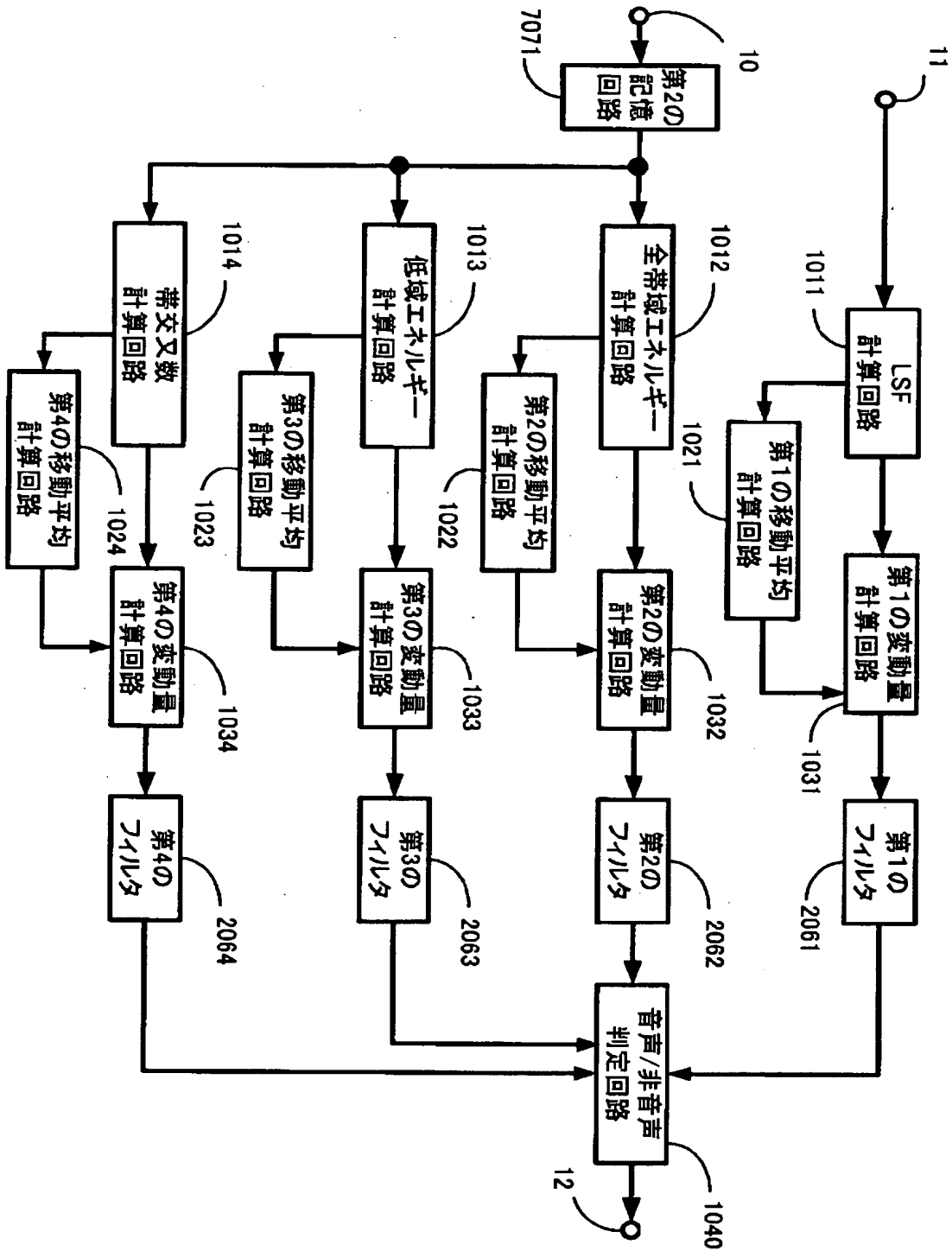
【図 1】



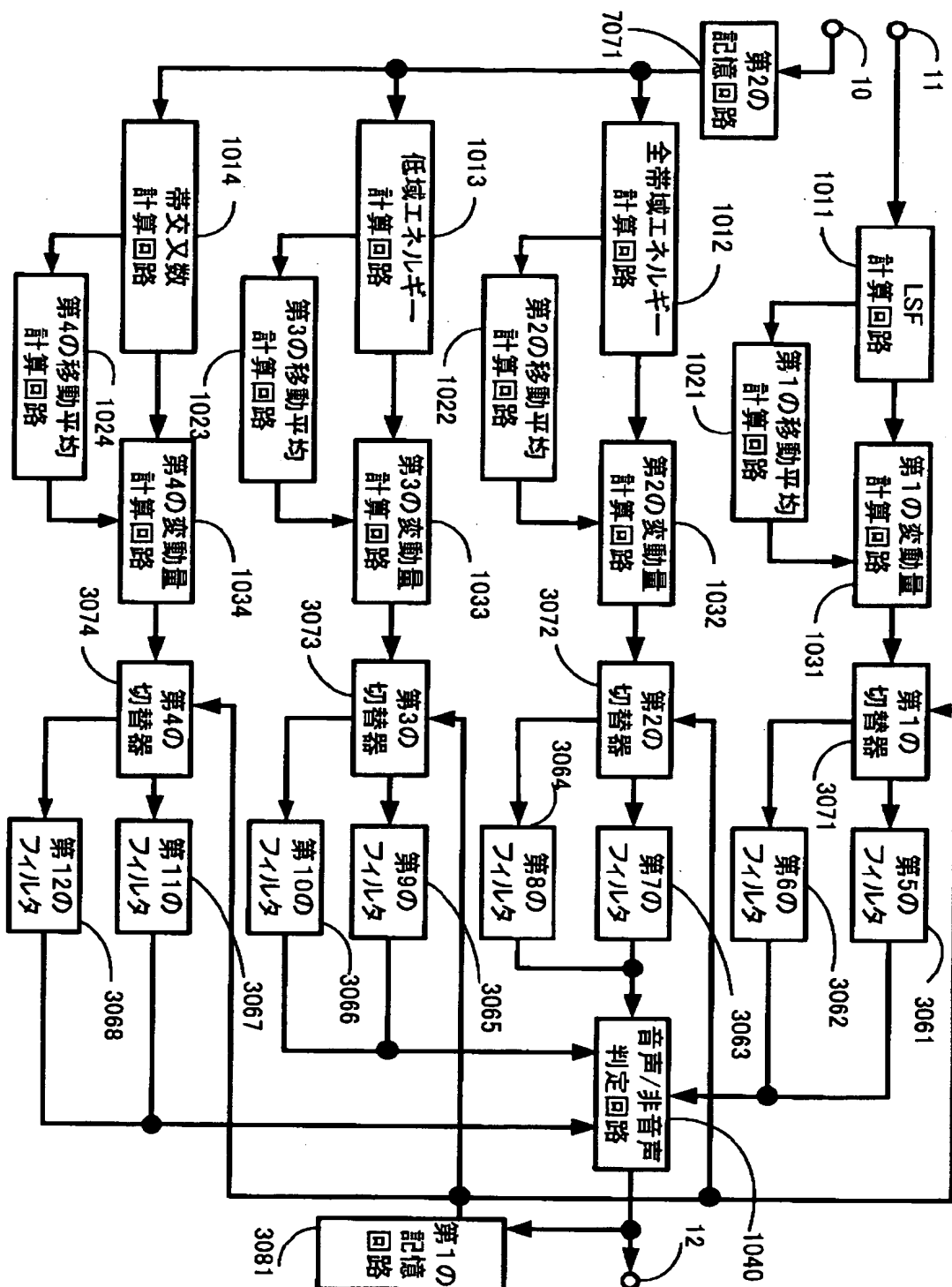
【図 2】



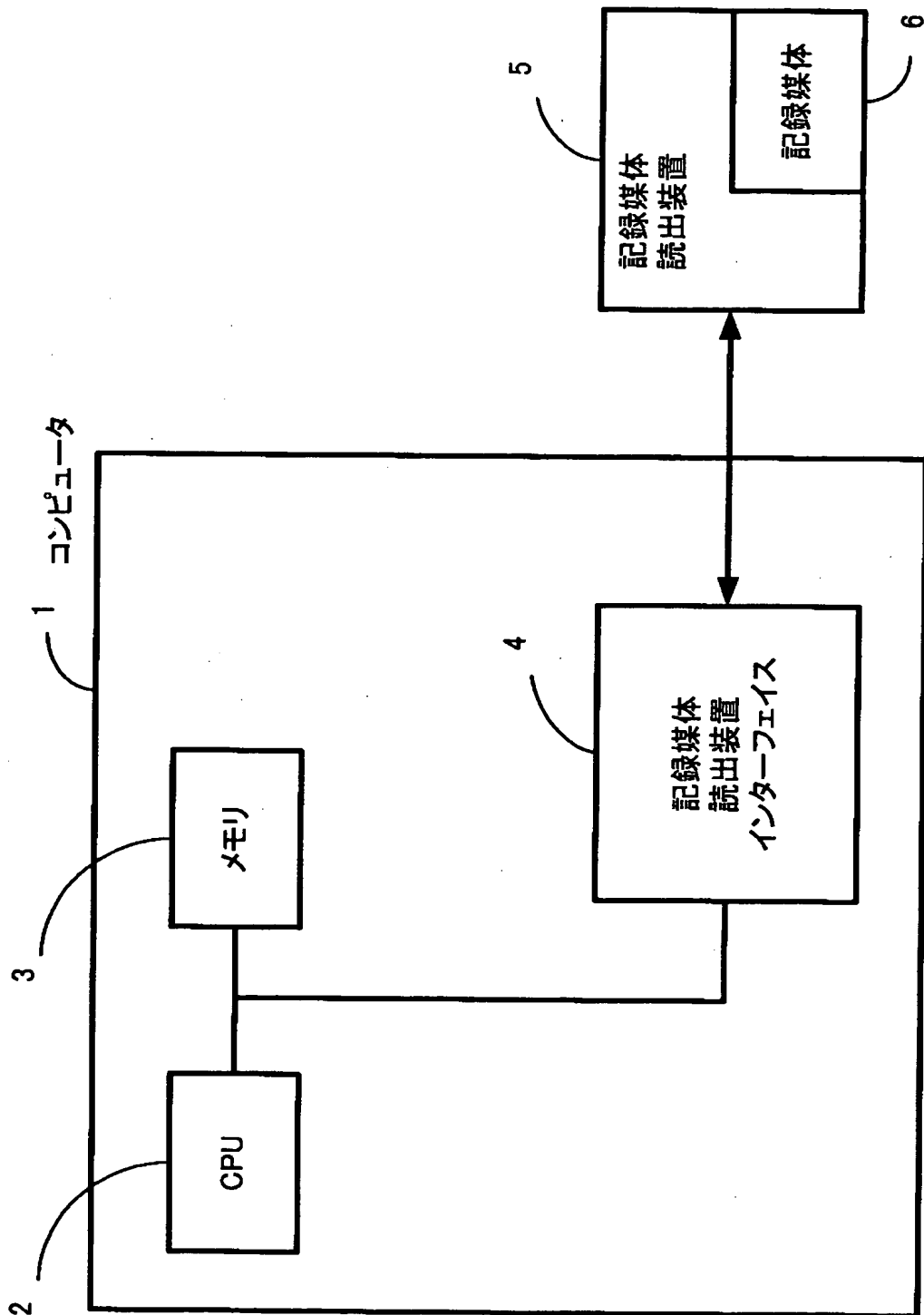
【図 3】

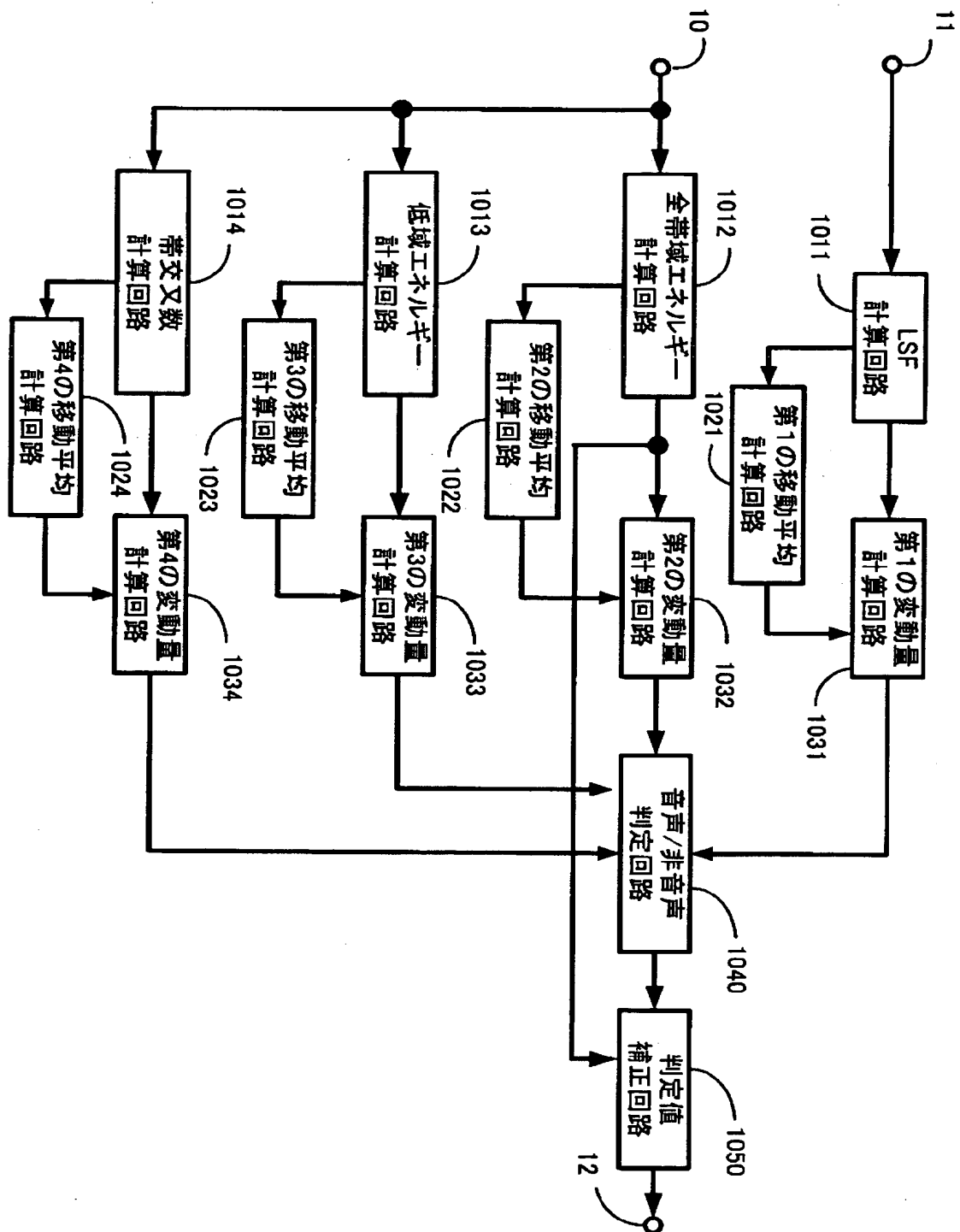


【図 4】

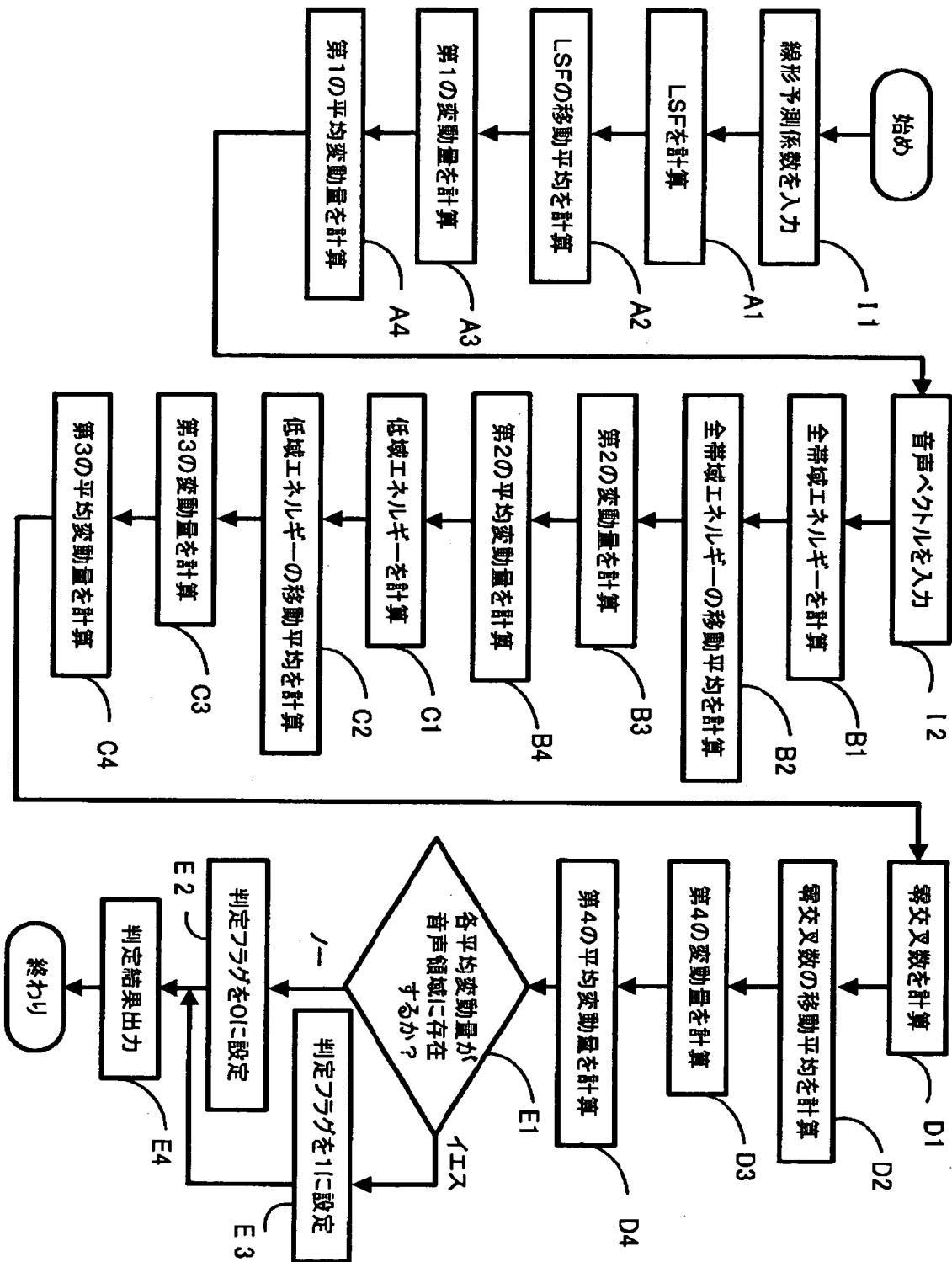


【図5】

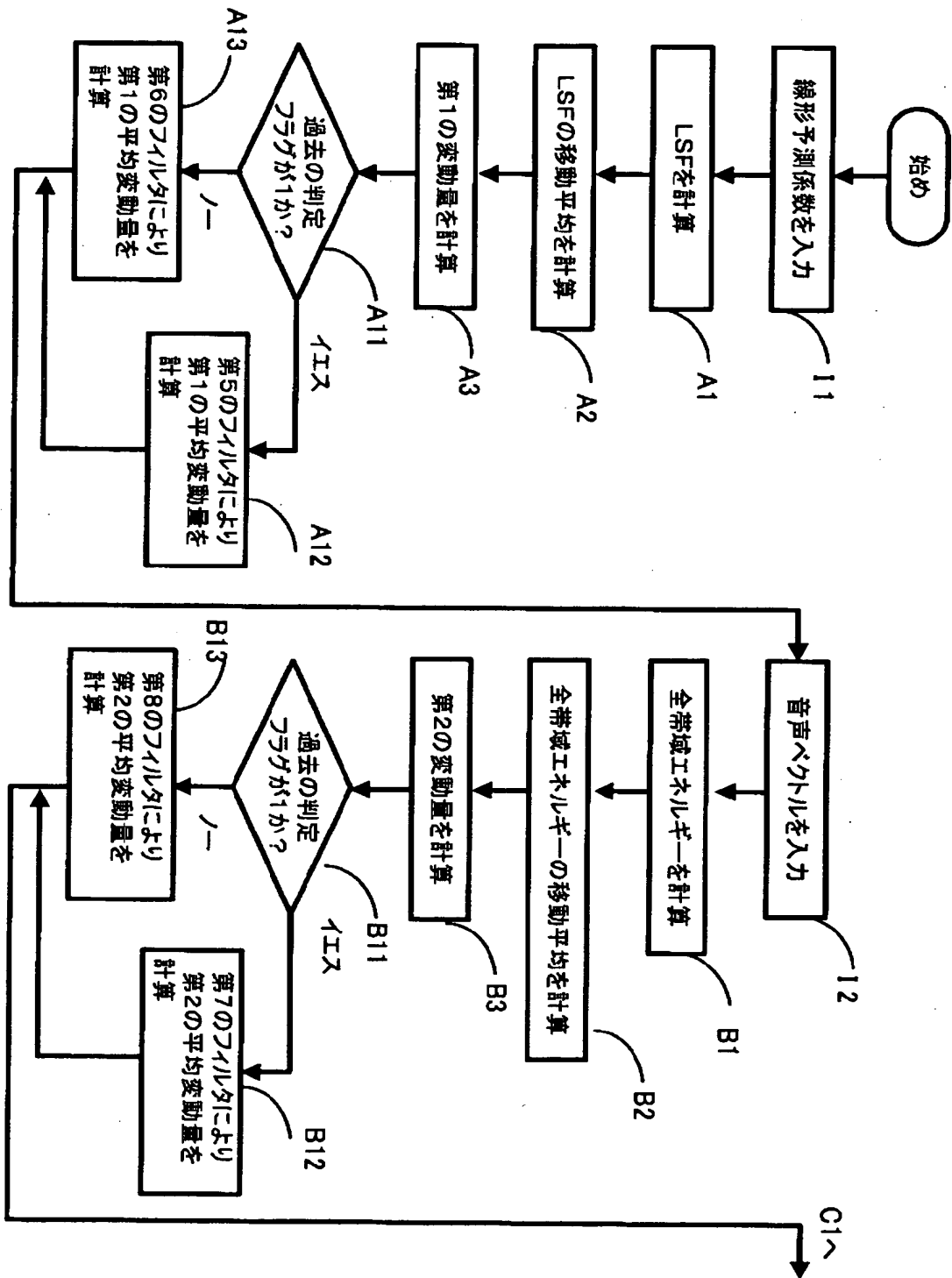




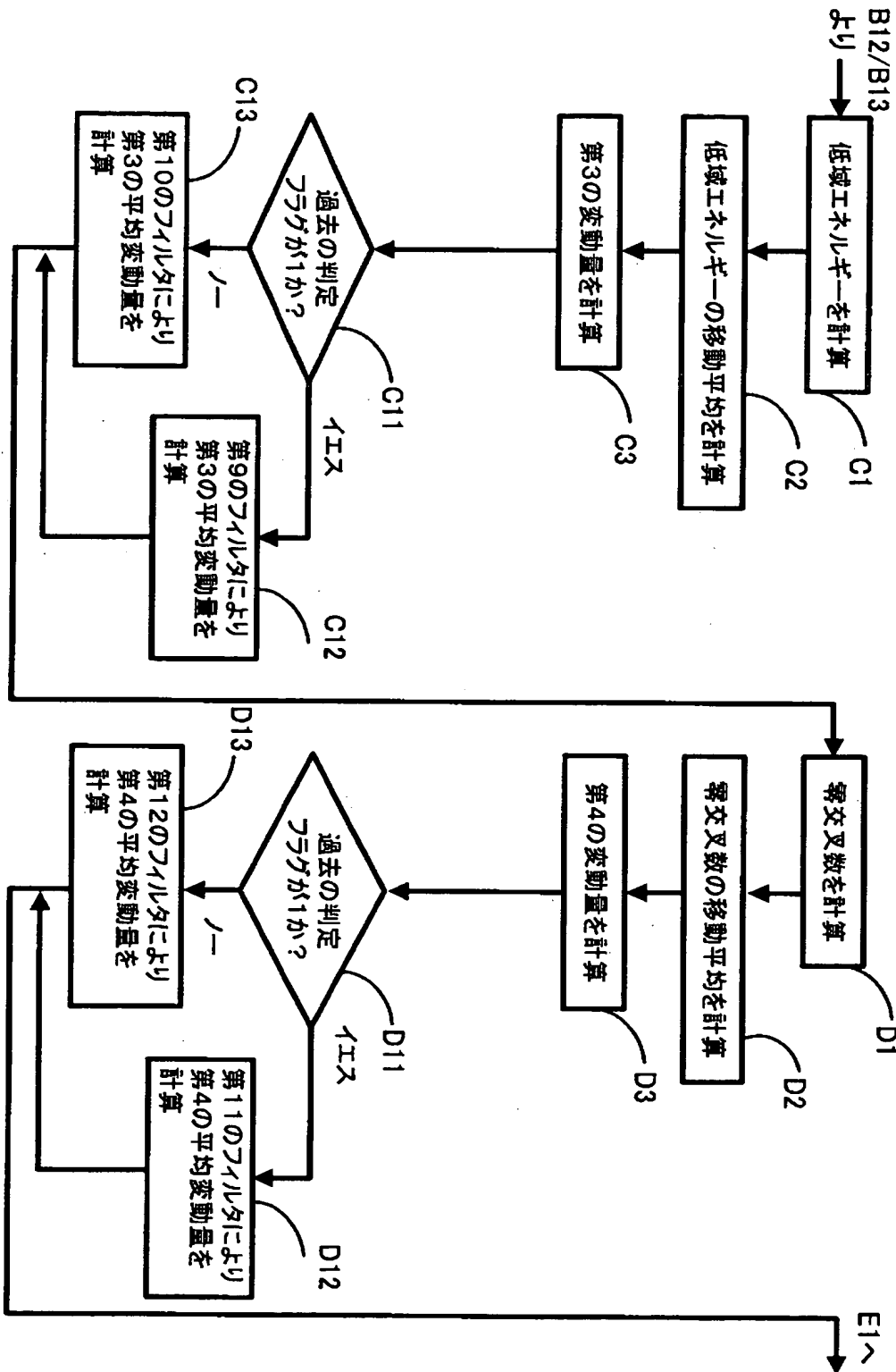
【図6】



【図7】

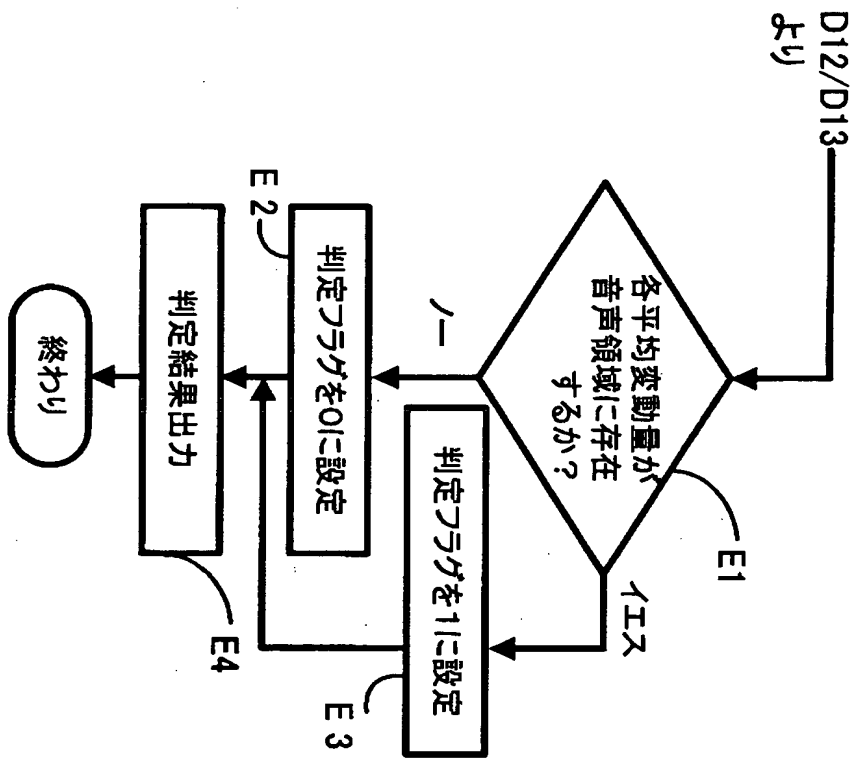


【図8】

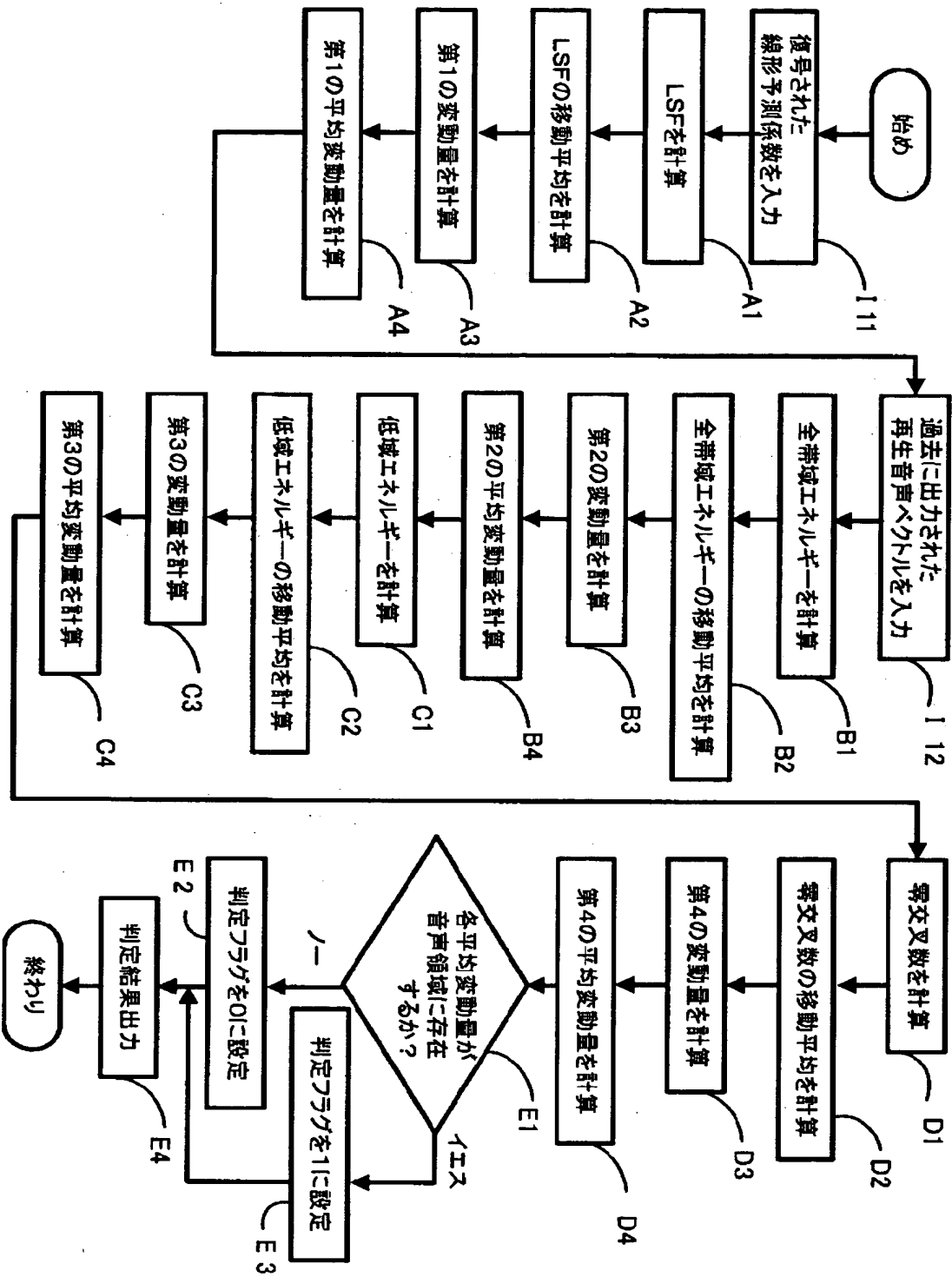


【図9】

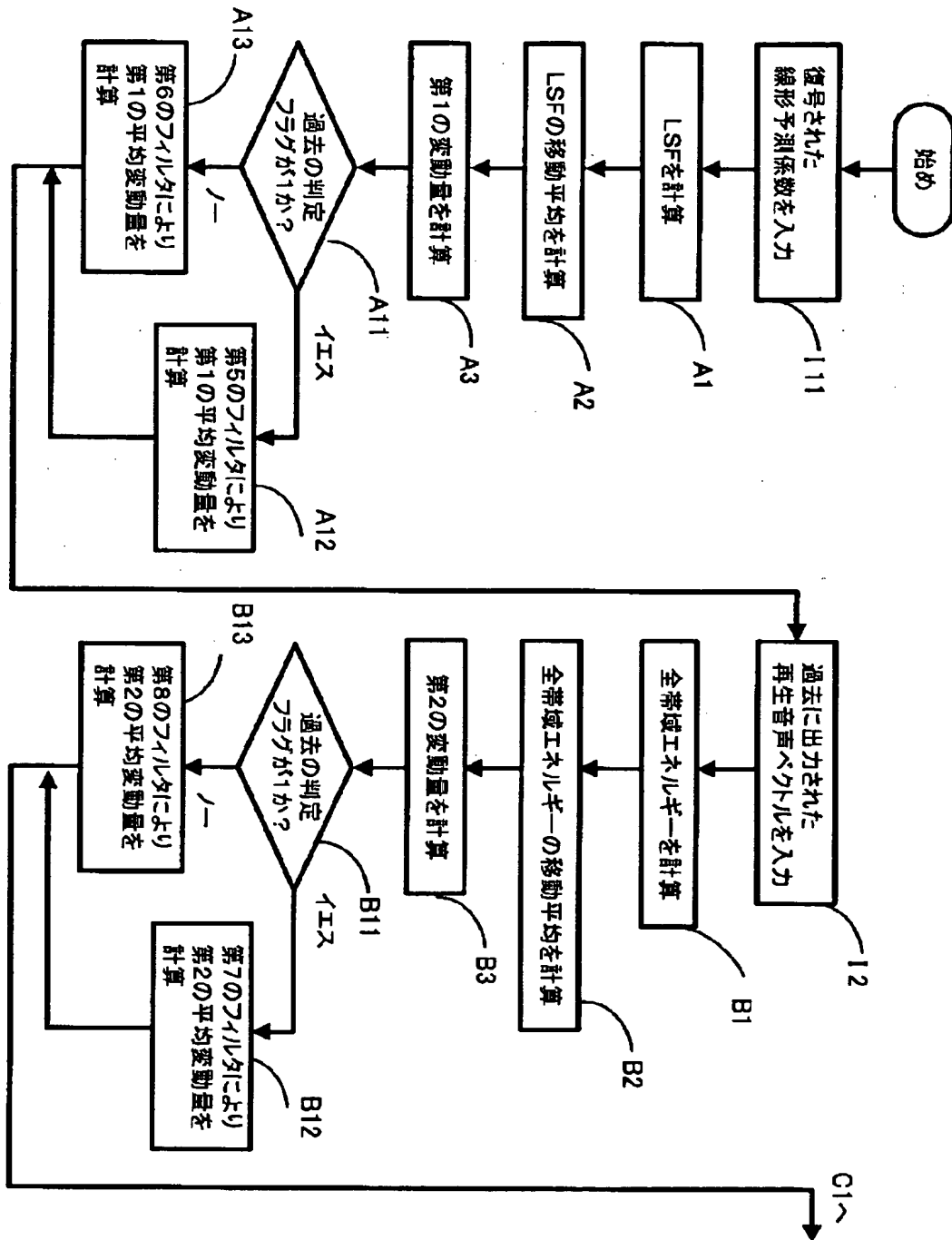
【図10】



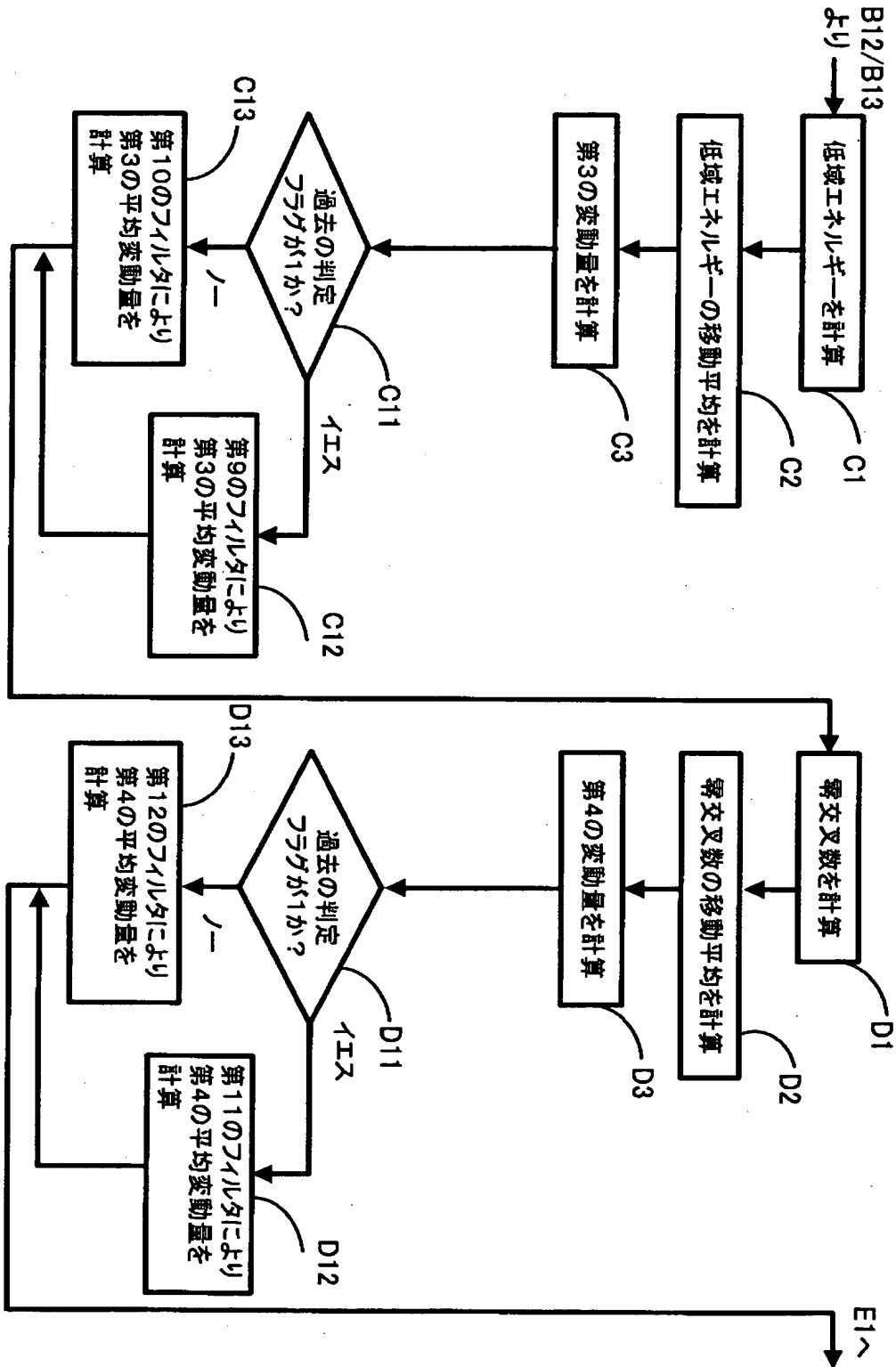
【図11】



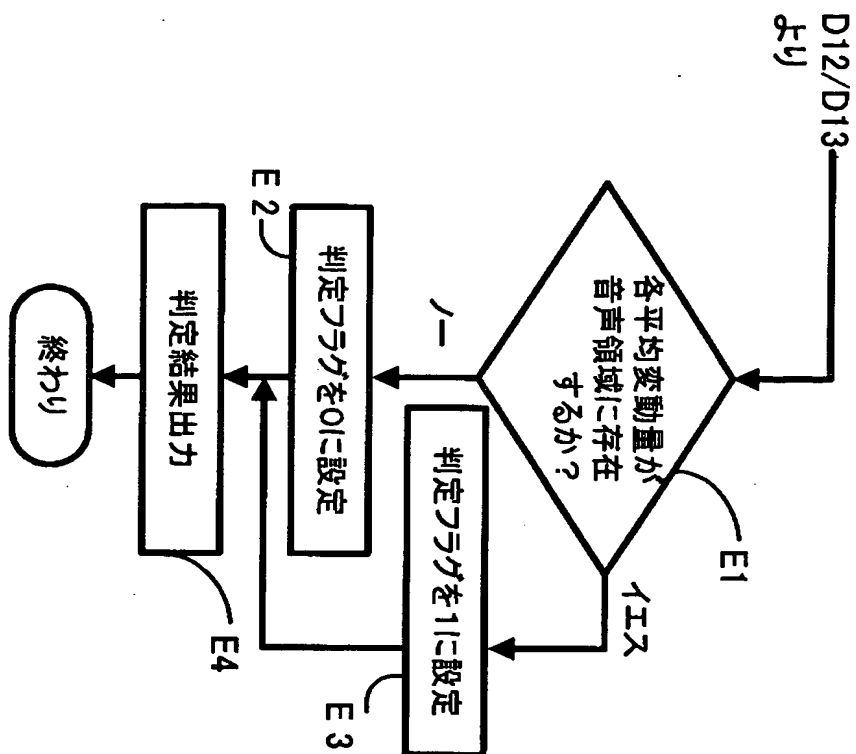
【図 12】



【図 13】



【図 1 4】



【書類名】 要約書

【要約】

【課題】 音声検出方法および装置の検出性能を改善する。

【解決手段】 第1のフィルタ（図1の2061）は、入力音声信号の線スペクトル周波数とその長時間平均との差分に基づく第1の変動量の長時間平均を計算する。第2のフィルタ（図1の2062）は、入力音声信号の全帯域エネルギーとその長時間平均との差分に基づく第2の変動量の長時間平均を計算する。第3のフィルタ（図1の2063）は、入力音声信号の低域エネルギーとその長時間平均との差分に基づく第3の変動量の長時間平均を計算する。第4のフィルタ（図1の2064）は、入力音声信号の零交叉数とその長時間平均との差分に基づく第4の変動量の長時間平均を計算する。音声/非音声判定回路（図1の1040）は、前記第1の変動量の長時間平均、前記第2の変動量の長時間平均、前記第3の変動量の長時間平均および前記第4の変動量の長時間平均を用いて音声信号を音声区間と非音声区間とに判別する。

【選択図】 図1

認 定 ・ 付 加 情 報

特許出願の番号	特願 2 0 0 0 - 1 6 6 7 4 6
受付番号	5 0 0 0 0 6 9 0 0 2 7
書類名	特許願
担当官	第八担当上席 0 0 9 7
作成日	平成 1 2 年 6 月 5 日

< 認定情報・付加情報 >

【提出日】	平成12年 6月 2日
-------	-------------

出 願 人 履 歴 情 報

識別番号 [000004237]

1. 変更年月日 1990年 8月29日
[変更理由] 新規登録
住 所 東京都港区芝五丁目7番1号
氏 名 日本電気株式会社